Die approbierte Originalversion dieser Diplom-/Masterarbeit ist in der Hauptbibliothek der Technischen Universität Wien aufgestellt und zugänglich.
http://www.ub.tuwien.ac.at

The approved original version of this diploma or master thesis is available at the main library of the Vienna University of Technology.
http://www.ub.tuwien.ac.at/eng

TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

D I P L O M A R B E I T

Paarhandel und Kointegration

Eine Machbarkeitsstudie anhand historischer DAX und MDAX Aktienkurse

Ausgeführt am Institut für

Stochastik und Wirtschaftsmathematik

der Technischen Universität Wien

unter der Anleitung von

Ao.Univ.Prof. Dr.techn. Wolfgang Scherrer

und der Mitbetreuung durch

Ao.Univ.Prof. Dr.techn. Friedrich Hubalek

durch

Peter Piekarz, BSc
Hauptstraße 29/4
A-2486 Landegg

Wien, 19. Oktober 2016 _____

Inhaltsverzeichnis

1 Einleitung **2**

2 Theoretische Grundlagen **5**

2.1 Grundlagen des Pairs-Trading 7

2.1.1 Der Distanz-Ansatz 10

2.1.2 Der Stochastische-Spread Ansatz 12

3 Integrierte und Kointegrierte Prozesse **15**

3.1 Das Granger Repräsentations-Theorem 16

3.2 Kointegration vs. Korrelation 19

3.3 Tests auf Kointegration 20

3.3.1 Das Engle-Granger 2-Schritt-Verfahren 21

3.3.2 Der Phillips-Ouliaris Test 23

3.3.3 Der Johansen-Test 26

3.4 Handelsstrategie bei Kointegration 30

4 Handelsstrategie **32**

4.1 Modellierungsphase 32

4.1.1 Auswahlkriterien 35

4.2 Handelsphase 38

4.2.1 Der Schwellwert „th“ 39

4.2.2 Die Hedge-Ratio γ 39

4.2.3 Sicherheitsmechanismen 41

4.2.4 Adaptive Schätzung mittels Kalman-Filter 42

5 Praxisteil **47**

5.1 Datenbeschreibung 47

5.1.1 Existieren Kointegrationsbeziehungen? 49

5.2 Parameter für die Handelsstrategie 52

5.2.1 Transaktionskosten 54

5.3 Resultate 55

6 Fazit **62**

Literaturverzeichnis **63**

1 Einleitung

Viele ökonomische Größen werden in regelmäßigen Abständen aufgezeichnet und mithilfe der Zeitreihenanalyse untersucht. Klassische Beispiele sind etwa Wetterdaten, Erhebungen von Arbeitslosenzahlen, die Inflationsrate oder Ergebnisse regelmäßig durchzuführender medizinischer Untersuchungen. Sehr oft denkt man jedoch bei Zeitreihen zunächst an Börsenkurse. Investoren sind seit jeher an der Entwicklung zuverlässiger Methoden zur Vorhersage von Aktienkursen interessiert, um mit möglichst wenig Risiko hohe Gewinne zu erzielen.
Existiert nun eine Investmentstrategie, die es risikolos ermöglicht Gewinn zu erwirtschaften, so spricht man von einer sogenannten (deterministischen) *Arbitrage-Möglichkeit*. In der Realität existieren solche Arbitrage-Möglichkeiten nur sehr selten und immer nur für sehr kurze Zeit - Investoren würden sofort versuchen, die Möglichkeit auf risikolosen Gewinn auszunützen und die dadurch bedingten Transaktionen würden die Preise der involvierten Finanzinstrumente derart anpassen, dass die Arbitrage-Möglichkeit verschwindet, vgl. Franke et al. (2008). Föllmer und Schied (2008) schreiben dazu:

"*The existence of such an arbitrage opportunity may be regarded as a market inefficiency in the sense that certain assets are not priced in a reasonable way. In real-world markets, arbitrage opportunities are rather hard to find. If such an opportunity would show up, it would generate a large demand, prices would adjust, and the opportunity would disappear.*"

Diese Diplomarbeit beschäftigt sich nicht mit deterministischen Arbitrage-Möglichkeiten, sondern mit sogenannten *statistischen* Arbitrage-Möglichkeiten. Darunter versteht man Investmentstrategien basierend auf statistischen Modellen, welche jedoch nicht risikolos sind. Meist wird unter statischer Arbitrage die Ausnutzung von „statistischen Fehlbewertungen" zweier Finanzinstrumente relativ zueinander verstanden. Dabei handelt es sich jedoch um eine auf einer historischen statistischen Beziehung basierende und keine tatsächliche Fehlbewertung der Preise. Da meist Paare von Aktien betrachtet werden, nennt man solche Strategien auch *Pairs-Trading*.
Die Anfänge des Pairs-Trading liegen in den 1980er-Jahre an der Wall Street. Nunzio Tartaglia und ein von ihm zusammengestelltes Team von Mathematikern, Physikern und Computerwissenschaftlern arbeiteten daran, Arbitrage-Strategien basierend auf modernsten statistischen Verfahren zu entwickeln. Eine dieser Strategien ist das

Pairs-Trading, dessen Grundidee sehr simpel ist: Es werden Paare von Wertpapieren gesucht, deren Preise - folgend einer statistischen Beziehung - sehr ähnlich verlaufen. Treten nun Abweichungen von dieser Beziehung auf - man vertraut jedoch auf ihre Wiederherstellung - dann kann daraus Gewinn erwirtschaftet werden, vgl. Vidyamurthy (2007).

In den folgenden Kapiteln werden zunächst einige Grundlagen zum Pairs-Trading erläutert, insbesondere werden drei unterschiedliche Ansätze vorgestellt - der *Distanz-Ansatz*, der *Stochastische-Spread Ansatz* und der *Kointegrations-Ansatz*. Heute wird insbesondere der Kointegrations-Ansatz mit statistischer Arbitrage in Verbindung gebracht, wobei Kointegration ein von Granger (1981) vorgestelltes Konzept ist, welches später von Engle und Granger (1987) weiter entwickelt wurde.

Anschließend an die Grundlagen des Pairs-Trading wird die Kointegrationstheorie genauer betrachtet. Kointegration beschreibt die Eigenschaft zweier oder mehrerer Zeitreihen, einen gemeinsamen stochastischen Trend zu besitzen. Betrachtet man zum Beispiel zwei kointegrierte Aktienkurse, dann bedeutet dies, dass die Preisdifferenz der beiden Aktienkurse zwar schwanken darf, jedoch nie wirklich divergieren wird. Tatsächlich schwankt eine Linearkombination der Preise um ein gewisses Gleichgewicht. Diese Eigenschaft - man nennt sie auch Mittelwertrückkehr - ist ausschlaggebend für die Entwicklung von Pairs-Trading Handelsstrategien basierend auf kointegrierten Aktienpaaren.

Zentraler Punkt dieses Abschnittes ist das *Granger-Repräsentationstheorem*, welches eine Zusammenfassung der Resultate von Engle und Granger (1987) darstellt. Basierend auf den Aussagen dieses Theorems lassen sich verschiedene Tests auf Kointegration herleiten. Vorgestellt werden zunächst der sehr intuitive Engle-Granger Test - auch bekannt als Engle-Granger 2-Schritt-Verfahren - und zwei Test-Statistiken nach Phillips und Ouliaris (1990). Hierbei handelt es sich jeweils um Kointegrationstests, bei denen die Residuen einer linearen Regression der zu betrachtenden Variablen untersucht werden. Anschließend wird der Johansen-Test vorgestellt, welcher im Gegensatz dazu von einem Vektorautoregressiven (VAR) Modelle dieser Variablen ausgeht, vgl. Johansen (1991).

Das folgende Kapitel beschäftigt sich mit Handelsstrategien basierend auf Kointegrationsmodellen. Ausgehend von einer „klassischen“ Strategie basierend auf einer konstanten Hedge-Ratio werden die Modellierungsphase und die Handelsphase genauer betrachtet und die notwendigen Parameter der Strategie werden erläutert. Außerdem sollen sogenannte Sicherheitsmechanismen besprochen werden, welche eventuelle Verluste beschränken sollen. Anschließend wird mittels Kalman-Filter eine adaptives Modell betrachtet, bei der die Kointegrationsparameter und somit die Hedge-Ratio nicht mehr als

konstant angenommen werden. Dadurch ergibt sich eine der im praktischen Teil dieser Arbeit getesteten „adaptiven“ Handelsstrategien
Der letzte Abschnitt ist der Praxisteil und somit Kern der vorliegenden Diplomarbeit. Ausgangspunkt der Arbeit war eine Machbarkeitsstudie zum Pairs-Trading anhand historischer DAX und MDAX Kurse, welche zunächst genauer beschrieben und in der freien Programmiersprache R auf Kointegrationsbeziehungen untersucht und auf Stabilität dieser Beziehungen getestet werden. Anschließend war es Ziel, profitable Handelsstrategien in R zu implementieren und anhand der vorliegenden historischen Aktienkurse zu testen. Es werden die Resultate von vier unterschiedlichen Handelsstrategien diskutiert - der klassischen Strategie mit konstanter Hedge-Ratio und dreier adaptiver Strategien - wobei jede dieser Strategien sowohl mittels des Engle-Granger Tests als auch mittels des Johansen Tests durchgeführt wurden.

2 Theoretische Grundlagen

Existiert eine herkömmliche bzw. *deterministische Arbitrage-Möglichkeit*, dann können Investoren risikolos und kostenfrei Gewinn erwirtschaften. Eine solche Arbitrage tritt zum Beispiel bei Kursunterschieden zum selben Zeitpunkt an unterschiedlichen Orten auf, da die Kurse bereits vor Eingehen des Handels bekannt sind und somit kein Risiko auf Verlust besteht. Im Gegensatz dazu beschreibt der Begriff *statistische Arbitrage* das Ausnutzen von statistisch begründeten Fehlbewertungen. Geht der Investor bei Erkennen einer solchen Kurs-Fehlbewertung Positionen ein, so geschieht dies nicht gänzlich ohne Risiko. Der erwartete Profit ist natürlich positiv, andernfalls würde wohl kaum gehandelt werden, ebenso ist jedoch auch die Wahrscheinlichkeit auf Verlust in diesem Fall positiv. Eine formale Definition für statistische Arbitrage liefern beispielsweise Hogan et. al. (2004) folgendermaßen:

Definition 1 (Statistische Arbitrage) *Es seien $(\xi_t)_{t\geq 0}$ eine selbstfinanzierende Handelsstrategie mit initialen Kosten von Null und $(v_t)_{t\geq 0}$ der zugehörige diskontierte Wertpapierprozess. Wir nennen ξ_t eine statistische Arbitrage-Möglichkeit, falls die folgenden Punkte erfüllt sind:*

$$\lim_{t\to\infty} \mathbb{E}[v_t] > 0 \tag{1}$$

$$\lim_{t\to\infty} \mathbb{P}[v_t < 0] = 0 \tag{2}$$

$$\lim_{t\to\infty} \frac{\mathbb{V}[v_t]}{t} = 0, \quad \textit{falls } \mathbb{P}[v_t < 0] > 0,\ \forall t < \infty \tag{3}$$

Eine statistische Arbitrage Möglichkeit hat nach dieser Definition im Grenzwert einen positiven erwarteten Gewinn (1) und die Wahrscheinlichkeit auf Verlust konvergiert gegen Null (2). Bedingung (3) sagt aus, dass die über die Zeit gemittelte Varianz des Wertprozesses gegen Null geht, falls die Wahrscheinlichkeit Verlust zu machen nicht in endlicher Zeit Null wird. Diese Bedingung ist somit nur notwendig, falls die Wahrscheinlichkeit auf Verlust immer positiv ist. Existiert hingegen ein endliches $N < \infty$, sodass $\mathbb{P}[v_t < 0] = 0$ für alle $t > N$ gilt, so existiert sogar eine deterministische Arbitrage-Möglichkeit, da ab diesem Zeitpunkt die Wahrscheinlichkeit auf Verlust gleich Null ist. Hogan et. al. zeigen auch, dass Bedingung (3) tatsächlich notwendig ist, da sonst im Black-Scholes Modell eine herkömmliche *Buy-and-Hold* Strategie - d.h. eine Wertanlage wird gekauft und langfristig gehalten - die Definition von statistischer Arbitrage erfüllen würde.

Statistische Arbitrage-Möglichkeiten basieren auf der relativen Fehlbewertung mehrerer Aktientitel, welche durch statistische Modelle erkannt und ausgenutzt werden soll. Eine der häufigsten Formen ist dabei der sogenannte Paar-Handel (auf Englisch *Pairs-Trading*), bei dem zwei Aktientitel als Paar relativ zueinander betrachtet werden. Der Grundgedanke ist es, die zu teure Aktie zu verkaufen und die zu günstige Aktie zu kaufen sobald sich deren Preise nicht im Gleichgewicht befinden. Gewinn wird bei Eingehen der genau entgegengesetzten Positionen gemacht, sobald sich die Preise erneut im Gleichgewicht befinden. Das Risiko auf Verlust liegt in der ungewissen Stabilität der Langzeit-Beziehung dieses Aktienpaares. Im nächsten Abschnitt werden wir uns mit dem Pairs-Trading beschäftigen und uns zwei solcher Strategien etwas genauer ansehen. Die Kointegrationstheorie, welche das Fundament dieser Diplomarbeit darstellt, wird im darauffolgenden Kapitel gründlicher bearbeitet.

2.1 Grundlagen des Pairs-Trading

Beim Pairs-Trading handelt es sich, wie der Name bereits andeutet, um Handelsstrategien basierend auf zwei unterschiedlicher Aktien, die als Paar betrachtet werden. Solche Strategien profitieren nicht direkt davon, dass die Kurse der Aktien steigen oder fallen, sondern von Fehlbewertungen der beiden Aktienkurse relativ zueinander. Stammen beispielsweise beide Aktien aus dem selben Industriesektor (etwa dem Energiesektor), dann verlaufen ihre Kurse oftmals sehr ähnlich, da sie den selben Markteinflüssen unterworfen sind. Genau diese „Ähnlichkeit“ soll beim Pairs-Trading ausgenutzt werden. Verlaufen nämlich die Preise eines solchen Aktienpaares sehr ähnlich, d.h. es lässt sich ein gewisses Gleichgewicht[1] feststellen, so sollen temporäre Abweichungen von diesem ausgenutzt werden. Dies geschieht durch simultanes (Leer-) Verkaufen der relativ zur anderen Aktie überbewerteten und Kaufen der relativ zur anderen Aktie unterbewerteten Aktie[2]. Anschließend wartet man auf den „richtigen“ Zeitpunkt, um beide Positionen glattzustellen[3]. Durch das Ausnutzen dieser kurzzeitigen Fehlbewertung der Aktienpreise wird somit Gewinn erwirtschaftet. Ein solcher „vollständig ausgeführter“ Handel - darunter verstehen wir das Eingehen der Long- und Short-Positionen bei Abweichung vom Gleichgewicht und deren Glattstellung bei der Rückkehr zum Gleichgewicht - wird als *Handelszyklus* bezeichnet. Ein großer Vorteil des Pairs-Tradings ist es, dass Investoren sowohl von fallenden als auch von steigenden Kursen profitieren können. Durch ein geeignet gewähltes Verhältnis der Anzahl an Long-Positionen und Short-Positionen ist es außerdem möglich, Markt-Neutrale Strategien zu konstruieren, vgl. etwa Lin et al (2006) bzw. Miao (2014).

Zunächst wollen wir klären, was wir unter dem „richtigen“ Zeitpunkt verstehen, an dem ein Handelszyklus wieder geschlossen wird. Die Fehlbewertung der Aktienpreise wird in den folgenden Kapiteln als stochastischer Prozess modelliert, welcher die Eigenschaft der „Mittelwertrückkehr“ besitzt:

[1] Eine genauere Definition dieses Gleichgewichts wird in den folgenden Abschnitten gegeben, da dieses vom verwendeten mathematischen Modell abhängt.

[2] Ab jetzt werden wir nur mehr von der über- und der unterbewerteten Aktie sprechen und meinen damit immer, dass dies relativ zueinander zu verstehen ist.

[3] Als Glattstellung bezeichnet man das Neutralisieren offener Positionen durch das Abschließen genau entgegengesetzter Positionen. Eine offene Long-Position wird zum Beispiel mit einer Short-Position zum selben Wert glattgestellt.

Definition 2 (Mittelwertrückkehr) *Wir bezeichnen einen stochastischen Prozess X_t mit endlichem Mittelwert $\mathbb{E}X_t = \mu < \infty$ mittelwertrückkehrend (auf Englisch "mean reverting"), falls er für alle $s \in \mathbb{N}$ und für alle $\epsilon > 0$ die folgenden Bedingungen erfüllt:*

$$\begin{aligned}\mathbb{P}[(X_t - \mu) > \epsilon,\ \forall t \geq s] = 0\\ \mathbb{P}[(X_t - \mu) < -\epsilon,\ \forall t \geq s] = 0.\end{aligned}$$

Bemerkung: Die Definition der Mittelwertrückkehr erfüllen beispielsweise Mittelwert-ergodische Prozesse, insbesondere ARMA-Prozesse. Dies lässt sich folgendermaßen einsehen:
Es sei X_t ein Mittelwert-ergodischer stochastischer Prozess mit endlichem Erwartungswert $\mu < \infty$, d.h. der Mittelwertschätzer

$$\bar{X}_T = \frac{1}{T}\sum_{i=1}^{T} X_i$$

konvergiert im quadratischen Mittel gegen den Erwartungswert μ. Es sei $\epsilon > 0$ gegeben, dann gilt:

$$\begin{aligned}0 &= \mathbb{P}[(X_t - \mu) \geq \epsilon,\ \forall t \geq 1]\\ &\leq \mathbb{P}[(X_t - \mu) \geq \epsilon,\ \forall t = 1, \ldots, T]\\ &\leq \mathbb{P}[\bar{X}_T - \mu \geq \epsilon]\\ &\leq \mathbb{P}[|\bar{X}_T - \mu| \geq \epsilon] \leq \frac{\mathbb{V}[\bar{X}_T]}{\epsilon^2}\end{aligned}$$

Der letzte Term konvergiert für $T \to \infty$ gegen Null, und da die Grenzwertbildung unabhängig von einer endlichen Anzahl an Folgengliedern ist folgt die erste Eigenschaft der Definition. Der Fall $(X_t - \mu) \leq -\epsilon$ wird analog behandelt.
Der „richtige" Zeitpunkt ist folglich genau dann, wenn die Fehlbewertung der Aktienpreise zum Gleichgewicht zurückgekehrt ist.

Beispiel 1 *Als Beispiel betrachten wir die Preise zweier fiktiver Aktien A und B, welche beide zur besseren Veranschaulichung zum Zeitpunkt $t = 0$ bei Eins starten.*

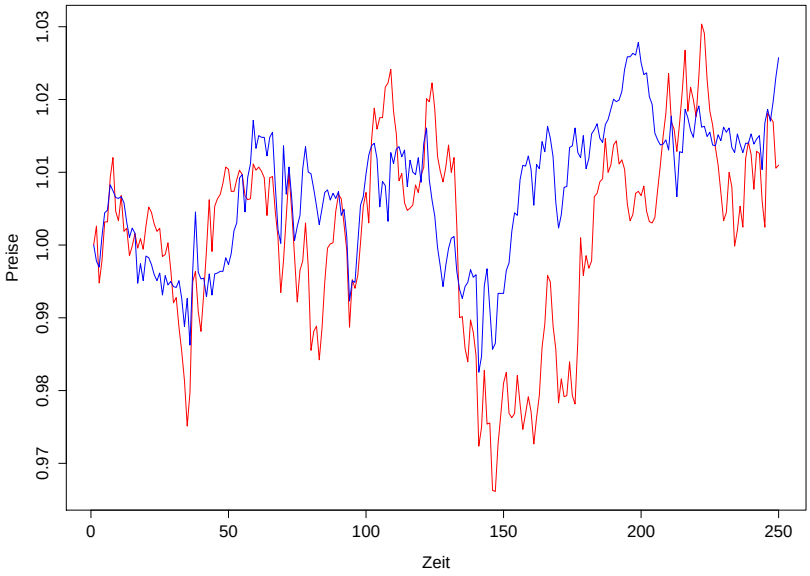


Abbildung 1: Die Preise der Aktien A und B.

Auch wenn beide Aktienkurse allem Anschein nach einem Random-Walk Prozess mit Drift folgen, so entfernen sie sich doch nie wirklich weit voneinander, sondern zirkulieren um besagtes Gleichgewicht. Noch klarer wird dies, falls man die Differenz

$$\epsilon_t := A - B$$

betrachtet, den sogenannten Spread[4] *dieses Aktienpaares:*

[4]Der Spread wird in den folgenden Kapiteln häufiger vorkommen, da er quasi als synthetisches Wertpapier verstanden und gehandelt werden kann. Dabei ist der Spread nicht immer von einer solch simplen Gestalt, er ist jedoch immer eine Linearkombination der beiden Aktientitel.

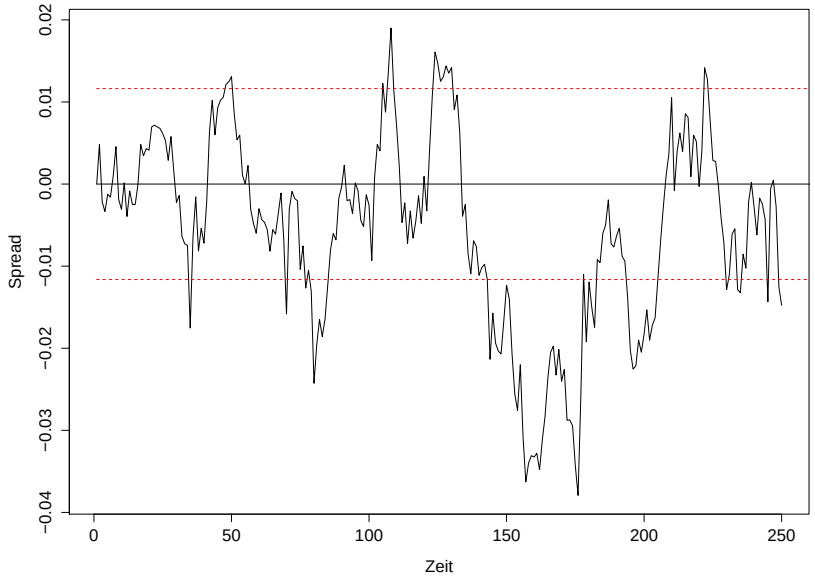


Abbildung 2: Der Spread zusammen mit seiner Standardabweichung (rot-strichliert).

Je weiter sich dieser Spread vom Gleichgewicht entfernt, desto stärker ist die relative Fehlbewertung der Preise und desto höhere Gewinne können bei einer Rückkehr zum Mittelwert erzielt werden. Long- und Short-Positionen werden üblicherweise erst eingegangen, falls sich der Spread hinreichend weit von seinem Mittelwert entfernt hat, beispielsweise ein Vielfaches seiner Standardabweichung. Einer der Gründe dafür ist, dass der Gewinn pro Handelszyklus mindestens die Transaktionskosten übersteigen sollte, da sonst insgesamt Verlust gemacht werden würde. Durchschreitet etwa der Spread einen positiven Schwellwert $th > 0$ *(in diesem Beispiel die einfache Standardabweichung), d.h. es gilt* $A - B \geq th > 0$*, dann ist der Aktientitel A relativ zum Aktientitel B überbewertet und/oder die Aktie B ist relativ zur Aktie A unterbewertet. Analog verhält es sich im Fall* $A - B \leq -th < 0$*.*

In den folgenden Abschnitten sollen zwei unterschiedliche Pairs-Trading Strategien kurz vorgestellt werden, da diese quasi als Vorläufer der statistischen Arbitrage - welche heute meist in Zusammenhang mit der Kointegrationstheorie betrachtet wird - gelten.

2.1.1 Der Distanz-Ansatz

Einer der historisch ältesten und zugleich intuitivsten Ansätze des Pairs-Trading ist die Beschreibung der Ähnlichkeit - d.h. eines möglichst „parallelen“ Verlaufes - zweier Aktienkurse durch deren Abstand, der *Distanz* (in englischsprachiger Literatur auch

oft als *Closeness* bezeichnet). Unter der Distanz der Aktienpreise A_t und B_t auf einem Zeitraum $t = 1, \ldots, T$ versteht man die Summe der quadrierten Abstände der normierten Preise[5],

$$D_{A_t,B_t} := \sum_{t=1}^{T} \left(\frac{A_t}{A_1} - \frac{B_t}{B_1} \right)^2 .$$

Die normierten Preise werden verwendet, da auch Aktien mit deutlichem Preisunterschied sehr ähnlich verlaufen können. Üblicherweise wird diese Summe für alle in Frage kommenden Aktienpaare gebildet und eine vom Investor abhängige Anzahl an Paaren mit möglichst kleinem Wert wird zum Handeln verwendet. Hierfür betrachtet man den Spread der beiden normierten Preis-Zeitreihen, welcher durch die Differenz $\tilde{A}_t - \tilde{B}_t$ definiert ist. Die Indikatoren für das Eingehen eines Handels sind durch

$$\begin{aligned} \tilde{A}_t - \tilde{B}_t \geq th \quad &\Rightarrow \text{Short-Position in A, Long-Position in B} \\ \tilde{A}_t - \tilde{B}_t \leq -th \quad &\Rightarrow \text{Long-Position in A, Short-Position in B} \end{aligned}$$

gegeben, wobei $th > 0$ ein meist von der historischen Standardabweichung des Spreads abhängiger Schwellwert ist. Ist der erste Fall eingetreten, d.h. zu einer Zeit t_0 ist der Spread größer oder gleich dem Schwellwert, so werden die offenen Positionen glattgestellt, falls der Spread erneut kleiner oder gleich Null ist. Der zweite Fall verhält sich analog dazu.

Zu den Vorteilen des Distanz-Ansatzes gehört, dass er eine parameterfreie Möglichkeit des Paar-Handels erlaubt und es somit zu keiner Fehlspezifikation kommen kann. Dadurch können jedoch auch keine Prognosen getroffen werden, wie beispielsweise der Stochastische-Spread Ansatz (vgl. Abschnitt 2.1.2) oder die Kointegrationstheorie (vgl. Kapitel 3) ermöglichen. Ein weiterer Nachteil ist die Annahme, dass die Differenz der Preis-Niveaus der beiden Aktientitel nahezu konstant bleibt. Dies ist jedoch meist nur für relativ kurze Zeit der Fall. Der Stochastische-Spread Ansatz, welcher im folgenden Abschnitt betrachtet wird, ist ein deutlich komplizierteres Konzept und wird vor allem im Zusammenhang mit adaptiver Schätzung (mittels Kalman-Filter) und dynamischer Modellierung verwendet.

[5] Die normierten Preise $\tilde{A}_t$ eines Aktienkurses A_t starten zum Zeitpunkt $t = 1$ bei Eins und weisen immer die selben Returns wie A_t auf, d.h. $\tilde{A}_t = \frac{A_t}{A_1}$.

2.1.2 Der Stochastische-Spread Ansatz

Beim Stochastischen-Spread Ansatz handelt es sich um eine Modellierung des Spreads zweier Aktientitel als gestörte Beobachtung eines versteckten, mittelwertrückkehrenden stochastischen Prozesses. Elliot et al. (2005), Do et al. (2006), Mudchanatongsuk et al. (2008) und Triantafyllopoulos und Montana (2009) beschreiben dieses Verfahren folgendermaßen:
Es sei $(X_t)_{t\in\mathbb{N}}$ ein reellwertiger stochastischer Prozess der zusätzlich nach Definition 2 mittelwertrückkehrend ist. Weiters nehmen wir an, dass X_t die folgende Darstellung

$$X_t - X_{t-1} = (a - bX_{t-1}) + \sigma\epsilon_t \tag{4}$$

besitzt, wobei $\sigma \geq 0$, $a \in \mathbb{R}$ sowie $b > 0$ konstante sind und $\epsilon_t \sim N(0,1)$ ein unabhängig identisch verteiltes weißes Rauschen beschreibt. Umformen von Gleichung (4) liefert den AR(1)-Prozess

$$X_t = a + (1-b)X_{t-1} + \sigma\epsilon_t,$$

welcher für $|1-b|< 1$ stationär ist und den Mittelwert $\mathbb{E}X_t = a/b$ besitzt.
Der beobachtbare Spread S_t ist definiert als die Differenz $A_t - B_t$ der beiden Aktientitel. Dieser Spread wird nun als Summe des Prozesses X_t und eines unabhängig identisch verteilten weißen Rauschens $\omega_t \sim N(0,1)$, welches unabhängig von ϵ_t ist, modelliert

$$S_t = X_t + D\omega_t \tag{5}$$
$$X_t = A + BX_{t-1} + C\epsilon_t, \tag{6}$$

wobei wir $A := a$, $B := 1-b$ und $C := \sigma$ gesetzt haben und $D > 0$ gilt. Zusammen mit der Zustands-Gleichung (5) beschreibt die Beobachtungs-Gleichung (6) ein lineares Zustandsraum-Modell mit den Parametern A, B, C, D.
Die dazugehörige Handelsstrategie basiert nun auf Schätzungen der bedingten Erwartungen

$$\hat{X}_{t|k} := \mathbb{E}[X_t|\mathcal{Y}_{k-1}] \tag{7}$$
$$\hat{X}_{t|k} := \mathbb{E}[X_t|\mathcal{Y}_k] \tag{8}$$

wobei $\mathcal{Y}_k := \sigma(S_0, \ldots, S_k)$ die von den Beobachtungen $S_0, \ldots, S_k$ aufgespannte Sigma-Algebra ist. Diese Erwartungswerte können mittels Kalman-Filter(vgl. Kapitel Kalman-

Filter) geschätzt werden, wobei der erste bedingte Erwartungswert einer Prognose des versteckten Zustandes X_t und der zweite Erwartungswert dem gefilterten Wert entspricht. Eine konkrete Handelsstrategie lässt sich nun folgendermaßen formulieren:
Ist der beobachtete Spread S_t zum Zeitpunkt t größer als seine Prognose aus der vergangenen Periode, d.h. $S_t > \hat{X}_{t|t-1} = \mathbb{E}[X_t|\mathcal{Y}_{t-1}]$, dann wird er als zu groß angenommen und ein Investor würde Long-Positionen in die Aktie A und Short-Positionen in die Aktie B eingehen. Üblicherweise geschieht dies erst, falls der Spread hinreichend viel größer als seine Prognose ist, beispielsweise $S_t \geq \hat{X}_{t|t-1} + th$ mit einem positiven Schwellwert $th > 0$. Glattgestellt werden die Positionen, falls sich eine Korrektur des Spreads eingestellt hat. Der Fall $S_t \leq \hat{X}_{t|t-1} - th$ wird analog dazu behandelt. Zusammengefasst ergibt dies die Handelsstrategie:

$$\begin{aligned} S_t &\geq \hat{X}_{t|t-1} + th \qquad &&\Rightarrow \text{Short-Position in A, Long-Position in B} \\ S_t &\leq \hat{X}_{t|t-1} - th \qquad &&\Rightarrow \text{Long-Position in A, Short-Position in B} \end{aligned}$$

Elliot et al. (2005) beschreiben zusätzlich eine weitere mögliche Handelsstrategie, welche auf einer Zeit-kontinuierlichen Approximation $\hat{X}_t$ der Zustandsvariablen X_t durch einen Ornstein-Uhlenbeck Prozess basiert. Diese erfüllt die stochastische Differentialgleichung

$$d\tilde{X}_t = b(\mu - \tilde{X}_t)dt + \sigma dW_t,$$

wobei W_t ein Wiener Prozess ist und die Variable b die „Geschwindigkeit“ - oft auch als Anziehungskraft bezeichnet - angibt, mit welcher sich der Prozess $\tilde{X}_t$ zu seinem Langzeitgleichgewicht μ hin korrigiert. Die Handelsstrategie nutzt nun bekannte Resultate zur ersten Eintrittszeit von Ornstein-Uhlenbeck Prozessen,

$$T := \inf(t \geq 0 : \tilde{X}_t = \mu | \tilde{X}_0 = \mu + \delta), \tag{9}$$

wobei $\delta > 0$ eine gewählte Konstante ist und als Schwellwert zum Öffnen eines Handels verstanden werden kann. Da Ornstein-Uhlenbeck Prozesse die Markov-Eigenschaft erfüllen, ist diese Eintrittszeit explizit zu berechnen. Elliot et al. (2005) schlagen vor, eine Long-Position in A und Short-Positionen in B einzugehen, falls $S_k \geq \mu + \delta$ gilt. Glattgestellt werden diese Positionen zum Zeitpunkt T später, wobei T die in (9) definierte erste Eintrittszeit ist. Der Fall $S_k \leq \mu - \delta$ verläuft analog.
Ein Vorteil des stochastischen Spread-Ansatzes gegenüber des Distanz-Ansatzes ist die konkrete Modellierung der Mittelwertrückkehr, welche ein essenzieller Bestandteil des

Paar-Handels ist. Außerdem sind die Modellparameter des Zustandsraum-Modells (5),(6) sehr einfach mittels Kalman-Filter zu schätzen. Weiters ist es durch die Verwendung der ersten Eintrittszeit der Ornstein-Uhlenbeck Approximation möglich, die erwartete Haltezeit (das ist jene Zeit, die zwischen Öffnen eines Handels und dessen Glattstellung vergeht) explizit zu berechnen. Do et al. (2006) weisen jedoch darauf hin, dass auch bei diesem Ansatz die Preis-Niveaus der beiden Aktien (oder deren Returns, falls der Spread als Differenz der logarithmierten Preise modelliert wird) als konstant angenommen werden.

3 Integrierte und Kointegrierte Prozesse

Der Begriff *Kointegration* geht auf die Nobelpreis-gewinnende Arbeit von Engle und Granger (1987) zurück. Kointegration beschreibt die Eigenschaft, dass zwei oder mehr stochastische Prozesse einen gemeinsamen stochastischen Trend besitzen, welcher sich durch eine geeignete Linearkombination eliminieren lässt. Für eine genauere Definition benötigen wir jedoch den Begriff der Integration:

Definition 3 *Ein stochastischer Prozess X_t, welcher nach mindestens d-maligem differenzieren eine stationäre und invertierbare ARMA-Darstellung besitzt, heißt integriert von der Ordnung d oder kurz $X_t \sim I(d)$.*
Besonders hervorzuheben sind die Fälle $d = 0$ und $d = 1$. Beim ersten Fall ist X_t selbst stationär und beim zweiten Fall ist der Prozess der absoluten Veränderungn $\Delta X_t = X_t - X_{t-1}$ stationär.

Ein Beispiel für integrierte Zeitreihen sind etwa Aktienkurse, welche selbst meist nicht stationär sind, jedoch durch Bilden der ersten Differenzen in eine stationäre Zeitreihe transformiert werden können. Daraus folgt, dass Aktienkurse häufig integriert von erster Ordnung sind. Bei dieser Transformation gehen allerdings Informationen verloren.
Eine beliebige Linearkombination mehrerer $I(d)$-Prozesse ergibt im Allgemeinen wieder einen $I(d)$-Prozess. Es ist jedoch möglich, dass sich die stochastischen Trends dieser Prozesse durch eine geeignete Linearkombination gegenseitig „abschwächen“ , d.h. der resultierende Prozess ist integriert von niedrigerer Ordnung, $I(\tilde{d})$ mit $\tilde{d} < d$. Speziell interessant ist der stationäre Fall $\tilde{d} = 0$, bei dem sich die Trends gegenseitig eliminieren. Betrachten wir hierzu das folgende formale Beispiel:

Beispiel 2 *Es seien die beiden stochastischen Prozesse X_t und Y_t durch*

$$X_t = \phi_X Z_t + \mu_t$$
$$Y_t = \phi_Y Z_t + \nu_t$$

gegeben, wobei Z_t ein $I(1)$-Prozess und $\mu_t \sim N(0,1)$ bzw. $\nu_t \sim N(0,1)$ zwei voneinander unabhängige White-Noise Prozesse sind. Der Prozess

$$\phi_Y X_t - \phi_X Y_t = \phi_Y \mu_t - \phi_X \nu_t$$

ist als Linearkombination stationärer Prozesse ebenfalls stationäre.

Das in diesem Beispiel beschriebene Verhalten lässt sich nun folgendermaßen formalisieren:

Definition 4 *Es seien $x_{1,t}, \ldots, x_{N,t} \sim I(d)$ integrierte Prozesse der Ordnung $d \neq 0$, d.h. jede Variable $x_{i,t}$, $i \in \{1, \ldots, N\}$, lässt sich durch d-maliges differenzieren in einen stationären Prozess transformieren. Existiert ein Vektor $(\beta_1, \ldots, \beta_N)$ dessen Einträge alle ungleich Null sind (der sogenannte Kointegrationsvektor[6]), sodass die Linearkombination*

$$(\beta_1, \ldots, \beta_n) \cdot (x_{1,t}, \ldots, x_{n,t})' = \beta_1 x_{1,t} + \beta_2 x_{2,t} + \ldots + \beta_n x_{n,t} \tag{10}$$

integriert von Ordnung $d - b$ mit $d \geq b > 0$ ist, dann nennt man die Variablen $x_{1,t}, \ldots, x_{n,t}$ kointegriert oder kurz $x_{1,t}, \ldots, x_{n,t} \sim CI(d, b)$. Existieren $r > 0$ linear unabhängige Kointegrationsvektoren, dann bezeichnen wir mit r den Rang dieser Kointegrationsbeziehung.

Das Besondere an einer existierenden Kointegration ist es, dass sich dadurch stabile Langzeitgleichgewichte bei nicht-stationären Variablen auffinden lassen, ohne vorher einen Informationsverlust in Kauf nehmen zu müssen (beispielsweise durch Differenzenbildung der Zeitreihen). Da Aktienkurse, wie bereits erwähnt, sehr häufig $I(1)$-Variablen sind, ist der für den Kontext von statistischer Arbitrage-Möglichkeiten relevante Fall ein Aktienpaar A und B mit $d = 1$ und $b = 1$. Die normierte Linearkombination aus Gleichung (10) für die zwei Aktienkurse A_t und B_t,

$$(1, -\beta) \cdot (A_t, B_t)' = A_t - \beta B_t$$

ist bei bestehender Kointegrationsbeziehung stationär und als ARMA-Prozess mittelwertrückkehrend. Diese Eigenschaft wird durch das Konzept der Kointegration - im Gegensatz zu beispielsweise dem Distanz-Ansatz - explizit modelliert.

3.1 Das Granger Repräsentations-Theorem

Engle und Granger (1987) leiten in ihrer Arbeit Resultate kointegrierter $I(1)$-Prozesse her, welche sie im sogenannten *Granger Repräsentations-Theorem* zusammenfassen. Dieses besagt beispielsweise, dass für Systeme kointegrierter $I(1)$-Variablen stets äquivalente

[6] Der Kointegrationsvektor ist nicht eindeutig bestimmt. Ist $(\beta_1, \ldots, \beta_N)$ ein Kointegrationsvektor, dann ist für jedes $\lambda \neq 0$ auch $(\lambda\beta_1, \ldots, \lambda\beta_N)$ ein Kointegrationsvektor. Üblicherweise wird durch die Wahl von $\lambda = 1/\beta_1$ der Kointegrationsvektor auf die erste Komponente normiert.

Darstellungen, unter anderem als *Fehlerkorrekturmodell* (auf Englisch *Error Correction Modell*), existieren. Als intuitiver Zugang zu dieser Darstellung soll das folgende motivierende Beispiel nach Engle und Granger (1987) dienen:

Beispiel 3 *Wir betrachten zunächst das Modell zweier Prozesse X_t und Y_t,*

$$Y_t + \alpha X_t = \nu_t, \qquad \nu_t = \nu_{t-1} + \xi_t \tag{11}$$

$$Y_t + \beta X_t = \epsilon_t, \qquad \epsilon_t = \rho\epsilon_{t-1} + \zeta_t, \quad |\rho| < 1 \tag{12}$$

wobei ξ_t und ζ_t zwei - nicht notwendigerweise voneinander unkorrelierte - unabhängig identisch verteilte White-Noise Prozesse und α bzw. β ungleich Null sind. Bei ν_t handelt es sich um einen Random-Walk Prozess - ν_t ist somit integriert von erster Ordnung - und ϵ_t ist ein stationärer Prozess, da $|\rho| < 1$ gilt. Aus den Gleichungen (11) *und* (12) *folgt, dass X_t und Y_t als Linearkombinationen von ν_t und ϵ_t jeweils $I(1)$-Prozesse sind und aus den Gleichungen* (12) *folgt die Stationarität der Linearkombination $Y_t + \beta X_t$. Die Prozesse X_t und Y_t sind folglich kointegriert mit Langzeitgleichgewicht $\mu = 0$ und Kointegrationsvektor $(1, \beta)'$. Durch Bilden der ersten Differenzen können die obigen Gleichungen umgeformt werden zu*

$$\Delta Y_t = \alpha\delta(Y_{t-1} + \beta X_{t-1}) + \eta_{1,t} \tag{13}$$

$$\Delta X_t = -\delta(Y_{t-1} + \beta X_{t-1}) + \eta_{2,t}, \tag{14}$$

wobei $\delta := (\rho - 1)/(\alpha - \beta)$ gilt und $\eta_{1,t}$ bzw. $\eta_{2,t}$ Linearkombinationen von ξ_t und ζ_t sind. In diesen Gleichungen ist $Y_{t-1} + \beta X_{t-1}$ der Störterm bzw. die Abweichung vom Langzeitgleichgewicht in der letzten Periode. Die jeweiligen Koeffizienten dieses Terms - d.h. $\alpha\delta$ bzw. $-\delta$ - beschreiben, wie sich das System zurück zu diesem Gleichgewicht korrigiert.
Lässt man den Parameter ρ gegen 1 konvergieren, so erhält man zwei korrelierte, jedoch nicht kointegrierte Random-Walk-Prozesse. Dies erkennt man auch in den Gleichungen (13) *und* (14)*, da δ dann gegen Null konvergiert.*

Die beiden Gleichungen (13) und (14) aus dem vorangegangenen Beispiel sind als Fehlerkorrekturmodell bekannt. Engle und Granger zeigen, dass für Systeme kointegrierter $I(1)$-Prozesse immer eine solche Darstellung existiert und weiters, dass diese sogar äquivalent zur Kointegrationsbeziehung ist. Zunächst wollen wir jedoch definieren, was diese Fehlerkorrektur-Darstellung eines multivariaten Systemes ist.

Definition 5 *Eine N-dimensionale multivariate Zeitreihe $X_t = (x_{1,t}, \ldots, x_{N,t})'$ besitzt eine Fehlerkorrektur-Darstellung, falls ein $N \times N$-Matrix-Polynom*

$$A(L) = I_N + \sum_{i=1}^{p} A_i L^i$$

des Lag-Operators L und $N \times r$-Matrizen α und β vollen Ranges r existieren, sodass sich der Differenzenprozess ΔX_t als

$$A(L)\Delta X_t = -\alpha Z_t + \epsilon_t$$

schreiben lässt, wobei wir $Z_t = \beta' X_t$ setzen und ϵ_t ein stationärer Prozess ist.

Das Granger-Repräsentationstheorem beinhaltet einige weitere nützliche Aussagen, welche wir nun zusammenfassen wollen, vgl. Engle und Granger (1987).

Theorem 1 (Granger-Repräsentationstheorem) *Es sei $X_t = (x_{1,t}, \ldots, x_{N,t})'$ ein N-dimensionaler Prozess integrierter Variablen erster Ordnung, d.h. für alle $i = 1, \ldots, N$ gilt $x_{i,t} \sim I(1)$. Weiters nehmen wir an, dass sein Differenzenprozess ΔX_t ein regulärer stationärer stochastischer Prozess mit Erwartungswert $\mathbb{E}\Delta X_t = 0$. Folglich besitzt ΔX_t die rein stochastische Woldsche Zerlegung*

$$\Delta X_t = (1 - L)X_t = \Phi(L)\epsilon_t$$
$$\Phi(L) = I_N + \Phi_1 L + \Phi_2 L^2 + \ldots$$

wobei $\Phi(L)$ ein $MA(\infty)$-Polynom des Lag-Operators und ϵ_t ein weißes Rauschen ist. Ist der Prozess $X_t \sim CI(1,1)$ kointegriert mit Kointegrationsrang $0 < r < N$, dann gelten folgende Aussagen:

(1) $\Phi(1)$ besitzt den Rang $N - r$

(2) Es existiert eine ARMA-Darstellung

$$A(L)X_t = B(L)\epsilon_t, \tag{15}$$

wobei $A(L) = I_N + \sum_{i=1}^{p} A_i L^i$ ein Polynom des Lag-Operators vom Grad p ist, $A(1)$ Rang r besitzt und $B(L)$ ein skalarwertiges Polynom des Lag-Operators.

(3) Es existieren $N \times r$-Matritzen α und β vollen Ranges r, sodass gilt:

$$\beta'\Phi(1) = 0, \qquad ,\Phi(1)\alpha = 0, \qquad A(1) = \alpha\beta'$$

(4) Es existiert eine Fehlerkorrektur-Darstellung mit dem Prozess $Z_t = \beta' X_t$,

$$A^*(L)\Delta X_t = -\alpha Z_{t-1} + B(L)\epsilon_t,$$

wobei $A^*(L)$ ein Polynom des Lag-Operators mit $A^*(0) = I_N$ und die Dimension von Z_t gleich r ist.

(5) Der Prozess $Z_t = \beta' X_t$ ist gegeben durch

$$Z_t = \beta'\Phi^*(L)\epsilon_t,$$

wobei $\Phi^*(L)$ ein Polynom des Lag-Operators ist und aus der Zerlegung $\Phi(L) = \Phi(1) + (1-L)\Phi^*(L)$[7] stammt.

3.2 Kointegration vs. Korrelation

Wie bereits in Beispiel (3) zu erkennen ist, sind die Begriffe Kointegration und Korrelation keines Falls synonym zueinander. Werden Zeitreihen auf eine Kointegrationsbeziehung untersucht, ist es daher äußerst wichtig, diese beiden Eigenschaften nicht zu verwechseln oder zu vermischen. Wird von korrelierten Aktienkursen A_t bzw. B_t gesprochen, so versteht man darunter, dass ihre Returns

$$r_t^A = \frac{A_t - A_{t-1}}{A_{t-1}}, \qquad r_t^B = \frac{B_t - B_{t-1}}{B_{t-1}}$$

miteinander korreliert sind. Steigt der Kurs der Aktie A, dann erwartet man sich bei einer stark-positiven Korrelation, dass auch der Kurs der Aktie B steigt. Bei kointegrierten Aktientiteln ist die Beziehung so zu verstehen, dass die beiden Aktien nicht zwingend in die selbe Richtung verlaufen müssen. Ein kointegriertes Aktienpaar entfernt sich jedoch nie sehr weit von seinem Langzeitgleichgewicht. Die Mittelwertrückkehr, eine Eigenschaft die Kointegration auszeichnet, muss bei lediglich korrelierten Aktienkursen nicht eintreten. Somit ist das Konzept der Kointegration wesentlich besser fürs Pairs-Trading geeignet als jenes der Korrelation.

[7] Eine solche Zerlegung existiert für alle Moving-Average Polynome Φ, vgl. Engle und Granger (1987).

3.3 Tests auf Kointegration

Eine existierende Kointegrationsbeziehung ist eine ausgezeichnete Grundlage für Investoren, um mittels Pairs-Trading Gewinn zu erwirtschaften. Dazu müssen solche Beziehungen jedoch erst nachgewiesen werden, was beispielsweise mit den in diesem Abschnitt vorgestellten Tests auf Kointegration erfolgen kann. Das folgende Beispiel zeigt, warum dies nicht *nur* durch eine simple OLS-Regression möglich ist.

$$A_t = \hat{\mu} + \hat{\beta} B_t + \hat{\epsilon}_t$$

geschehen kann.

Beispiel 4 (Spurious Regression) *Zwei kointegrierte Zeitreihen der Ordnung Eins, X_t und Y_t, erfüllen nach Definition 4 eine lineare Beziehung*

$$Y_t - \beta X_t = \epsilon_t,$$

wobei ϵ_t ein stationärer Prozess ist. Häufig liefert jedoch auch eine Regression voneinander unabhängiger Zeitreihen hohe R^2 und signifikante t-Werte, welche auf einen linearen Zusammenhang der Daten deuten, welcher tatsächlich jedoch nicht existiert. Granger und Newbold (1974) haben diesen als Scheinregression (auf Englisch Spurious Regression) bekannten Effekt erstmals untersucht. In ihrer Arbeit simulieren sie zwei voneinander unabhängige Random-Walk Prozesse X_t und Y_t. Anschließend schätzen sie die Koeffizienten β_0 und β_1 der Regressionsgleichung

$$Y_t = \beta_0 + \beta_1 X_t + \epsilon_t \tag{16}$$

mittels OLS-Schätzer. Der t-Test zum 5 %-Niveau lehnt fälschlicherweise die Nullhypothese $H_0 : \beta_1 = 0$ in ca. 75 % aller Fälle ab (Fehler 1. Art), bei Verwendung mehrerer Regressoren steigt die Ablehnungsrate sogar noch weiter an. Aus ihren Ergebnissen folgern Granger und Newbold, dass im Falle stark autokorrelierter Residuen - beispielsweise durch einen niedrige Durbin-Watson Statistik[8] indiziert - nur auf eine Fehlspezifikation des Modells geschlossen werden kann.
Dies folgt daraus, dass in der Beziehung (16) $\beta_1 = 0$ gilt und somit ϵ_t integriert von erster Ordnung und kein White-Noise Prozess ist. Der t-Test ist demnach nicht anwendbar. Die

[8] Der Durbin-Watson Test ist ein statistischer Test auf Autokorrelation erster Ordnung von Regressions-Residuen. Die Nullhypothese lautet, dass der Korrelationskoeffizient der Autokorrelation erster Ordnung gleich Null ist, d.h. $H_0 : \rho_1 = 0$.

(lineare) Beziehung zwischen zwei integrierten Prozessen kann folglich nicht nur durch OLS-Schätzung und mittels t-Test nachgewiesen werden. Hierfür sind speziell konstruierte Tests auf Kointegration notwendig.

Zunächst wird ein sehr intuitiver Test besprochen, welcher sich direkt durch die Definition von Kointegration und Beispiel 3 motivieren lässt - das sogenannte Engle-Granger 2-Schritt-Verfahren, vgl. Engle und Granger (1987). Dieser Test basiert ebenso wie die beiden anschließend vorgestellten Teststatistiken nach Phillips und Ouliaris (1990) auf Regressions-Residuen, wie sie im vorangegangenen Beispiel erwähnt werden. Zuletzt wird mit dem Johansen-Test ein Verfahren präsentiert, welches von einer VAR(p)-Darstellung eines Vektors integrierter Variablen selben Grades ausgeht und nicht direkt die Residuen einer linearen Regression betrachtet. Mittels zweier von Johansen (1988) vorgeschlagener Statistiken wird anschließend der Kointegrationsrang bestimmt.

3.3.1 Das Engle-Granger 2-Schritt-Verfahren

Engle und Granger liefern in ihrer Arbeit nicht nur eine Definition für den Begriff Kointegration, sondern sie schlagen auch einen sehr einfachen und intuitiven Test auf diese Eigenschaft vor. Die Motivation für diese als Engle-Granger Zwei-Schritt-Verfahren bekannte Testprozedur liefert bereits die Definition von Kointegration, denn laut dieser muss eine schwach-stationäre Linearkombination existieren.
Es sei $X_t = (x_{1,t}, \ldots, x_{N,t})'$ ein Vektor integrierter Variablen der Ordnung Eins, welche zusätzlich eine Kointegrationsbeziehung erfüllen. Nach Definition (4) existiert somit eine schwach-stationäre Linearkombination

$$\beta' X_t = \mu + \epsilon_t,$$

wobei $\beta = (1, \beta_2, \ldots, \beta_N)'$ ein normierter Kointegrationsvektor, μ eine Konstante und ϵ_t ein stationärer Prozess mit Erwartungswert Null sind. Die Linearkombination $\beta' X_t$ zirkuliert folglich um das Langzeit-Gleichgewicht μ. Bevor wir die beiden Schritte dieses Verfahrens durchführen muss mittels eines Einheitswurzeltestes - beispielsweise dem Dickey-Fuller Test - gewährleistet werden, dass alle Komponenten von X_t integriert der Ordnung Eins sind. Die beiden Schritte sind dann folgendermaßen gegeben:

Schritt 1: Der erste Schritt des Testverfahrens besteht nun aus einer linearen OLS-Regression einer der Variablen auf die andere. O.B.d.A. sei diese durch

$$x_{1,t} = \hat{\mu} + (\hat{\beta}_2, \ldots, \hat{\beta}_N) \cdot (x_{2,t}, \ldots, x_{N,t})' + \hat{\epsilon}_t$$

gegeben, wobei der Vektor $(\hat{\beta}_2, \ldots, \hat{\beta}_N)$ die geschätzten Regressions-Parameter und $\hat{\epsilon}_t$ die geschätzte Abweichung vom geschätzten Langzeit-Gleichgewicht $\hat{\mu}$ sind.

Schritt 2: Da ϵ_t stationär sein muss, ist der zweite Schritt des Engle-Granger Verfahrens das Überprüfen der Regressions-Residuen $\hat{\epsilon}_t$ auf Stationarität, etwa mittels des augmentierten Dickey-Fuller (ADF) Tests. Wichtig zu Bemerken ist, dass die aus der Regression erhaltenen Residuen nicht die tatsächlichen, sondern lediglich eine Schätzung sind, welche einer anderen Verteilung folgen. Die Parameter $(\hat{\beta}_2, \ldots, \hat{\beta}_N)$ und die Konstante $\hat{\mu}$ werden bei einer OLS-Regression so geschätzt, dass sie die Varianz des Fehlers $\hat{\epsilon}_t$ minimieren. Das bedeutet, dass die OLS-Schätzung möglichst stationär wirkende Residuen liefert, selbst für nicht kointegrierte Zeitreihen. Daher ist es nicht möglich, die regulären kritischen Werte des ADF-Tests zu verwendet. Durch Monte Carlo Simulationen berechnete kritsche Werte liefert zum Beispiel MacKinnon (1992, 2010).
Der große Vorteil des 2-Schritt Verfahrens nach Engle und Granger gegenüber anderen Kointegrationstests ist der sehr intuitive Aufbau und die relativ simple Implementierung. Ein Nachteil dieses Testes ist es jedoch, dass nur eine einzige Gleichung und somit auch nur eine Kointegrationsbeziehung gleichzeitig untersucht werden kann. Bereits beim bivariaten Fall - beispielsweise der fürs Pairs-Trading interessante Fall eines Aktienpaares - ist das Resultat der Teststatistik abhängig davon, welche Aktie als abhängige und welche als unabhängige Variable in der (OLS-)Regression verwendet wird. Untersucht man die beiden Aktienkurse A_t und B_t auf Kointegration, sowohl mit A_t als auch mit B_t als abhängiger Variable, d.h.

$$\begin{aligned} A_t &= \hat{\beta}_1 B_t + \hat{\mu}_1 + \hat{\epsilon}_{1,t} \\ B_t &= \hat{\beta}_2 A_t + \hat{\mu}_2 + \hat{\epsilon}_{2,t}, \end{aligned}$$

dann sind die beiden Steigungen $\hat{\beta}_1$ und $\hat{\beta}_2$ der Regressionsgeraden nicht invers zueinander, $\hat{\beta}_1 \neq \hat{\beta}_2^{-1}$. Tests auf Stationarität der Residuen $\hat{\epsilon}_{1,t}$ und $\hat{\epsilon}_{2,t}$ liefern in der Theorie asymptotisch die selben Ergebnisse, in der Praxis - d.h. einer endlichen Datenmenge - ist dies jedoch nicht der Fall.

3.3.2 Der Phillips-Ouliaris Test

Phillips und Ouliaris (1990) schlagen zwei weitere Kointegrationstests vor - die Varianz-Ratio Statistik $\hat{P}_u$ und die multivariate Spur Statistik (auf Englisch *multivariate trace statistic*) $\hat{P}_z$. Beide Tests gehen wie das Engle-Granger Verfahren von einer Regressions-Gleichung aus: Es seien y_t eine skalarwertige integrierte Variable vom Grad Eins sowie x_t ein (N-1)-dimensionaler Vektor integrierter Variablen vom Grad Eins. Die Regressions-Gleichung ist nun

$$y_t = \hat{b}' \cdot x_t + \hat{u}_t, \tag{17}$$

wobei $\hat{u}_t$ die geschätzten Residuen sind. Schreiben wir $Z_t := (y_t, x_t')$ - d.h. Z_t ist ein N-dimensionaler Vektor von I(1)-Variablen - dann entspricht $\hat{\beta} := (1, -\hat{b}')$ einem geschätzten Kointegrationsvektor, falls $\hat{u}_t$ stationär ist. Im Gegensatz zum Engle-Granger Verfahren werden die Regressions-Residuen $\hat{u}_t$ jedoch nicht direkt auf Stationarität überprüft. Zu Bemerken ist, dass es wie beim Engle-Granger Verfahren ebenfalls möglich ist, eine Regression

$$y_t = \hat{\mu} + \hat{b}' \cdot x_t + \hat{u}_t$$

mit konstantem Term $\hat{\mu}$ zu betrachten. Die Interpretationen der beiden Statistiken ändert sich dadurch nicht, die kritischen Werte müssen jedoch angepasst werden.
Eine Motivation für die Teststatistiken $\hat{P}_u$ und $\hat{P}_z$ lässt sich durch die folgende Feststellung liefern: Der Differenzenprozess ΔZ_t ist laut Voraussetzung stationär und besitzt eine MA(∞)-Darstellung

$$\Delta Z_t = \Psi(L)\epsilon_t, \tag{18}$$

wobei $\Psi(L) = \sum_{j\geq 0} \psi_j L^j$ ein Matrix-Polynom des Lag-Operators L und ϵ_t ein White-Noise Prozess mit Kovarianzmatrix Σ sind. Schreiben wir

$$\Psi^*(L) := (1-L)^{-1}[\Psi(L) - \Psi(1)],$$

dann erhalten wir die folgende Darstellung für Z_t

$$Z_t = \mu^* + \Psi(1)\sum_{j=1}^{t} \epsilon_t + \Psi^*(L)\epsilon_t,$$

wobei μ^* ein konstanter Vektor ist, vgl. Campbell und Perron (1991). Existiert nun ein Kointegrationsvektor β, so dass $\beta' Z_t$ stationär ist, dann muss folglich auch die rechte Seite der Gleichung

$$\beta' Z_t = \beta' \mu^* + \beta' \Psi(1) \sum_{j=1}^{t} \epsilon_t + \beta' \Psi^*(L) \epsilon_t,$$

stationär sein. Der erste Term ist konstant und der dritte Term ist stationär, da $\Psi(L)\epsilon_t = \Delta Z_t$ stationär ist. Der zweite Term beinhaltet jedoch den Random-Walk Prozess $\sum_{j=1}^{t} \epsilon_t$, womit $\beta'\Psi(1) = 0$ eine notwendige (und auch hinreichende) Bedingung für die Stationarität von $\beta' Z_t$ ist. Da β als Kointegrationsvektor sinnvollerweise nicht gleich dem Nullvektor ist, muss die Matrix $\Psi(1)$ singulär sein, d.h. von nicht vollem Rang $N - r$. Die Zahl $r > 0$ ist die Anzahl linear abhängiger Spalten von $\Psi(1)$ und folglich die Anzahl existierender, linear unabhängiger Vektoren β mit $\beta'\Psi(1) = 0$. Somit entspricht r gleichzeitig dem Kointegrationsrang.
Folgend der Notation von Phillips und Ouliaris (1990) schreiben wir den Differenzenprozess ΔZ_t nun als

$$\xi_t := \Delta Z_t = \Psi(L)\epsilon_t, \tag{19}$$

wobei vor allem die spektrale Dichte[9]

$$f_\xi(\omega) = \frac{1}{2\pi} \Psi(e^{-i\omega}) \; \Sigma \; \overline{\Psi(e^{-i\omega})}',$$

ausgewertet bei Frequenz ω gleich Null interessant ist:

$$f_\xi(0) = \frac{1}{2\pi} \Psi(1) \; \Sigma \; \Psi(1)'.$$

Für alle Kointegrationsvektoren β folgt aus $\beta'\Psi(1) = 0$, dass auch $\beta' f_\xi(0) = 0$ gilt und der Rang von $f_\xi(0)$ ebenfalls $N - r$ sein muss. Die spektrale Dichte ausgewertet bei der Frequenz Null entspricht bis auf den Faktor 2π der Langzeit-Kovarianz-Matrix Ω,

$$2\pi f_\xi(0) = \Omega := \sum_{j \in \mathbb{Z}} \gamma(h).$$

[9]Die diskrete Fouriertransformation $f_x(\omega) = 1/(2\pi) \sum_{h \in \mathbb{Z}} \gamma(h) e^{-i\omega h}$ der Autokovarianzfunktion $\gamma(h)$ eines stochastischen Prozesses x_t wird als spektrale Dichte von x_t bezeichnet.

Beide vorgeschlagenen Test-Statistiken basieren auf Schätzungen dieser Matrix Ω, da deren Singularität äquivalent zur Existenz einer Kointegrationsbeziehung ist. Angepasst an die Partitionierung $Z_t = (y_t, x_t')$ partitionieren Phillips und Ouliaris (1990) die Matrix Ω als

$$\Omega = \begin{bmatrix} \omega_{11} & \omega_{21}' \\ \omega_{21} & \Omega_{22} \end{bmatrix}.$$

Existiert keine Kointegrationsbeziehung der Variablen x_t - d.h. es existiert kein $(N-1)$-dimensionaler Vektor b, so dass $b'x_t$ stationär ist - dann ist die Matrix Ω_{22} regulär. In diesem Fall lässt sich zeigen, dass die Determinante von Ω durch

$$\det \Omega = \omega_{11.2} \cdot \det \Omega_{22}$$

gegeben ist, wobei $\omega_{11.2}$ durch $\omega_{11.2} := \omega_{11} - \omega_{21}' \Omega_{22}^{-1} \omega_{21}$ definiert ist. Sind die Variablen x_t nicht kointegriert, dann ist die Existenz einer Kointegrationsbeziehung der Variablen (y_t, x_t') äquivalent zu $\omega_{11.2} = 0$. Phillips und Ouliaris (1990) formulieren die Nullhypothese keiner vorliegenden Kointegrationsbeziehung und ihre Alternativhypothese durch

$$H_0 : \omega_{11.2} \neq 0 \quad vs. \quad H_A : \omega_{11.2} = 0$$

Eine Schätzung $\hat{\Omega}$ der Matrix Ω erhält man aus

$$\hat{\Omega} = T^{-1} \sum_{t=1}^{T} \hat{\xi}_t \hat{\xi}_t' + T^{-1} \sum_{s=1}^{L} g_{sL} \sum_{t=1}^{T} (\hat{\xi}_t \hat{\xi}_{t-s}' + \hat{\xi}_{t-s} \hat{\xi}_t'),$$

wobei $g_{sL} = 1 - \frac{s}{L+1}$ eine Gewichtungsfunktion mit gewähltem Lag L ist. Beide Statistiken $\hat{P}_u$ und $\hat{P}_z$ basieren auf den soeben motivierten Ergebnissen, anstelle der Darstellung (18) wird jedoch eine Regression

$$Z_t = \hat{\Pi} Z_{t-1} + \hat{\xi}_t, \tag{20}$$

betrachtet, wobei $\hat{\Pi}$ eine $N \times N$ Koeffizientenmatrix und $\hat{\xi}_t$ die Residuen dieser Regression sind. Werden die Residuen $\xi_t = \Delta Z_t$ aus Darstellung (19) zum Schätzen der Matrix Ω verwendet, dann bezeichnen Phillips und Ouliaris (1990) die Test-Statistiken mit P_u und P_z. Sie merken an, dass diese Unterscheidung wesentlich ist, da diese Statistiken unter

der Alternativhypothese einer vorliegenden Kointegrationsbeziehung nicht divergieren und somit inkonsistente Tests ergeben.

Varianz-Ratio Statistik Die Varianz-Ratio Statistik $\hat{P}_u$ ist definiert durch

$$\hat{P}_u = \frac{T\omega_{11.2}}{T^{-1}\sum_{t=1}^{T}\hat{u}_t^2}.$$

Unter der Alternativhypothese einer vorliegenden Kointegrationsbeziehung divergiert die Varianz-Ratio Statistik. Eine Tabelle passender kritischer Werte findet man bei Phillips und Ouliaris (1990).

Multivariate-Trace-Statistik Die *multivariate trace statistic* $\hat{P}_z$ ist durch

$$\hat{P}_z = T\mathrm{tr}\left(\hat{\Omega}M_{ZZ}^{-1}\right)$$

gegeben, wobei $M_{ZZ} = T^{-1}\sum_{t=1}^{T} Z_t Z_t'$ die empirische Momenten-Matrix ist. Der große Vorteil der $\hat{P}_z$-Statistik gegenüber der $\hat{P}_u$-Statistik, ist die Invarianz bezüglich der Reihung der Variablen. Die $\hat{P}_z$-Statistik liefert demnach immer das selbe Ergebnis, egal welche der Variablen von $Z_t = (y_t, x_t')$ als endogene und welche als exogene Variablen in der Regression (17) verwendet werden. Tabellen passender kritischer Werte finden sich ebenfalls bei Phillips und Ouliaris (1990).

3.3.3 Der Johansen-Test

Der letzte Kointegrationstest, der in dieser Arbeit besprochen werden soll, ist der von Johansen (1988, 1991) vorgeschlagene Johansen-Kointegrationstest. Dieser verwendet - anders als das Engle-Granger Verfahren und die beiden Phillips-Ouliaris Tests - keine Regression der Zeitreihen, um die dadurch geschätzten Residuen auf Stationarität zu überprüfen. Stattdessen handelt es sich hierbei um statistische Tests zur Schätzung des Kointegrationsranges r.
Es sei $X_t = (x_{1,t}, \ldots, x_{N,t})'$ ein N-dimensionaler Vektor integrierter Variablen der Ord-

nung Eins[10]. Der Johansen-Test geht nun von einer VAR(p)-Darstellung

$$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + \epsilon_t \tag{21}$$

aus, wobei ϕ_i $N \times N$-Koeffizientenmatrizen sind. Es spielt theoretisch keine Rolle, ob der Vektor X_t stationär ist oder integrierte Komponente besitzt, ohne jedoch kointegriert zu sein, da der Johansen-Test den Kointegrationsrang r untersucht, welcher im stationären Fall gleich N und im integriert jedoch nicht kointegrierten Fall gleich Null ist. Existiert für X_t jedoch eine Kointegrationsbeziehung, so nimmt r Werte zwischen Null und N an. Aus der VAR(p)-Darstellung (21) kommt man durch subtrahieren von X_{t-1} auf beiden Seiten und anschließendes Umstrukturieren zur folgenden formalen Darstellung des Differenzenprozesses ΔX_t

$$\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta X_{t-i} + \epsilon_t, \tag{22}$$

wobei $\Pi = \sum_{i=1}^{p} \phi_i - I$ und $\Gamma_i = -\sum_{j=i+1}^{p} \phi_j$ gilt. Bei dieser Darstellung handelt es sich um ein Fehlerkorrekturmodell wie es das Granger-Repräsentationstheorem postuliert. Da X_t laut Annahme aus integrierten Zeitreihen der Ordnung Eins besteht ist ΔX_t stationär. Folglich muss auch der Term ΠX_{t-1} stationär sein, damit die obige Gleichung erfüllt sein kann. Für die Matrix Π existieren nun drei mögliche Fälle betreffend ihres Ranges:

(1) rang $\Pi = 0$, d.h. Π ist die Nullmatrix. Das Fehlerkorrekturmodell vereinfacht sich zu einem VAR(p-1) Prozess der ersten Differenzen von X_t, d.h.

$$\Delta X_t = \sum_{i=1}^{p-1} \Gamma_i \Delta X_{t-i} + \epsilon_t$$

Der Prozess X_t hat somit N Einheitswurzeln und es existiert keine Kointegrationsbeziehung der $x_{i,t}$.

(2) rang $\Pi = N$, d.h. Π hat vollen Rang und ist invertierbar. Da jede Linearkombina-

[10]Es ist auch Möglich, stationäre Variablen in X_t aufzunehmen. Die Matrix Π aus Gleichung (22) - welche Ausgangspunkt der Untersuchungen auf Kointegration ist - hat dann jedoch immer einen reduzierten Rang. Wir beschränken uns deshalb darauf, dass alle Komponente von X_t integriert der Ordnung Eins sind.

tion stationärer Variablen ebenfalls stationär ist, folgt aus

$$\Pi^{-1}\Delta X_t = X_{t-1} + \Pi^{-1}\epsilon_t,$$

dass X_{t-1} stationär sein muss, d.h. X_t besteht zur Gänze aus stationären Variablen. In diesem Fall existiert ebenfalls keine Kointegrationsbeziehung der $x_{i,t}$.

(3) $0 <$ rang $\Pi = r < N$, d.h. Π ist eine singuläre Matrix ungleich der Nullmatrix. Die Koeffizientenmatrix Π lässt sich nun in das Produkt zweier $N \times r$ Matrizen α und β vollen Ranges r zerlegen, d.h. es gilt $\Pi = \alpha\beta'$. In dieser Zerlegung entsprechen die Spalten von β den r linear unabhängigen Kointegrationsvektoren - $\beta' X_t$ ist somit stationär - und die Elemente der Matrix α sind die Koeffizienten des Fehlerkorrekturmodelles und geben an, wie schnell die einzelnen Variablen sich dem Langzeitgleichgewicht annähern.

Johansen (1991) schlägt zwei unterschiedliche Tests vor, um die Matrix Π auf ihren Rang hin zu untersuchen. Zunächst werden Eigenwerte einer Matrix $\hat{S}$ geschätzt, wobei jeder signifikant von Null verschiedene Eigenwert einem existierenden Kointegrationsvektor entspricht. Da Matrix $\hat{S}$ ist so konstruiert, dass es sich bei ihren Eigenwerten um die Quadrate sogenannter kanonischer Korrelationen handelt und somit alle Eigenwerte zwischen Null und Eins liegen. Bezeichnen wir mit $1 \geq \tilde{\lambda}_1 \geq \tilde{\lambda}_2 \geq \ldots \geq \tilde{\lambda}_N \geq 0$ die absteigend geordneten Eigenwerte dieser Matrix, dann ergeben sich die folgenden beiden Tests:

Trace-Statistik Die Trace-Statistik testet die Nullhypothese, dass genau $r_0 < N$ linear unabhängige Kointegrationsvektoren existieren gegen die Alternativhypothese, dass mindestens $r_0 + 1$ linear unabhängige Kointegrationsvektoren existieren:

$$H_0 : r = r_0 \quad vs. \quad H_A : r > r_0$$

Die Trace-Statistik ist dann durch

$$\hat{\lambda}_{trace} = -T \sum_{i=r_0+1}^{N} \ln(1 - \lambda_i)$$

gegeben, wobei $T > 0$ die Anzahl der Stichproben von X_t bezeichnet.

Maximum-Eigenvalue-Statistik Die Maximum-Eigenvalue-Statistik testet die Nullhypothese, dass genau $r_0 < N$ linear unabhängige Kointegrationsvektoren existieren gegen die Alternativhypothese, dass genau $r_0 + 1$ linear unabhängige Kointegrationsvektoren existieren:

$$H_0 : r = r_0 \quad vs. \quad H_A : r = r_0 + 1$$

Die Maximum-Eigenvalue-Statistik ist dann durch

$$\hat{\lambda}_{max} = -T \ln(1 - \lambda_{r_0+1})$$

gegeben, wobei $T > 0$ die Anzahl der Stichproben von X_t bezeichnet.
Beide Test-Statistiken basieren darauf, dass im Falle einer singulären $N \times N$-Matrix von Rang $0 < r < N$ exakt r verschiedene Eigenwerte unterschiedlich von Null sind. Somit sollten beide Test-Statistiken bei zutreffender Nullhypothese kleine Werte liefern. Eine Tabelle kritischer Werte findet man etwa in Johansen und Juselis (1990).
Anzumerken ist, dass es sich bei beiden Tests um sogenannte *nested* Tests handelt. Daher sollte immer mit der Nullhypothese „keine Kointegration“ , $H_{0,1} : r = 0$, begonnen werden. Wird diese Nullhypothese zugunsten ihrer Alternativ-Hypothese „es existiert mindestens eine Kointegrationsbeziehung“ , $H_{A,1} : r \geq 1$, verworfen, so ist die nächste zu testende Nullhypothese durch $H_{0,2} : r = 1$ gegeben.

3.4 Handelsstrategie bei Kointegration

Wie sieht nun eine mögliche Pairs-Trading Handelsstrategie bei vorliegender Kointegrationsbeziehung aus? Wie beim Distanz-Ansatz und beim Stochastischen-Spread Ansatz soll eine Abweichung vom Langzeitgleichgewicht der Preise ausgenutzt werden. Betrachten wir hierzu ein als kointegriert klassifiziertes Aktienpaar A und B, welches die Regressionsgleichung

$$A_t = \hat{\mu} + \hat{\beta} B_t + \hat{\epsilon}_t$$

erfüllt. In dieser Gleichung entspricht $(1, -\hat{\beta})$ dem geschätzten Kointegrationsvektor und $\hat{\epsilon}_t$ ist die geschätzte Abweichung vom geschätzten Langzeitgleichgewicht $\hat{\mu}$. Als Spread bezeichnen wir nicht wie bei den zuletzt besprochenen Ansätzen die Differenz $A_t - B_t$, sondern die Linearkombination

$$A_t - \hat{\beta} B_t = \hat{\mu} + \hat{\epsilon}_t,$$

welche aufgrund der Klassifikation als kointegriertes Aktienpaar schwach-stationär ist. Analog zu den bereits besprochenen Pairs-Trading Strategien ergibt sich dadurch folgende Strategie

$$\begin{aligned} A_t - \hat{\beta} B_t \leq \hat{\mu} - th \qquad & \Rightarrow \text{Short-Position in A, Long-Position in B} \\ A_t - \hat{\beta} B_t \geq \hat{\mu} + th \qquad & \Rightarrow \text{Long-Position in A, Short-Position in B} \end{aligned}$$

wobei $th > 0$ wieder ein positiver Schwellwert ist. Üblicherweise wird die Anzahl der Positionen die Long bzw. Short gegangen werden durch die sogenannte *Hedge-Ratio* γ festgelegt. Diese Ratio gibt an, wie viele Positionen in die Aktie B eingegangen werden sollen für jede Position die in die Aktie A eingegangen wird. Ist beispielsweise der Spread größer als das Gleichgewicht $\hat{\mu}$, dann werden γ Aktien B gekauft für jede verkaufte Aktie A. Üblicherweise wird als Hedge-Ratio der geschätzte Kointegrationskoeffizient $\hat{\beta}$ verwendet, was sich unmittelbar durch die Kointegrationsbeziehung motivieren lässt. Außerdem lässt sich dadurch unter Vernachlässigung eventueller Transaktionskosten ein minimaler Profit von $th > 0$ erzielen: Gilt für den Spread zu einem Zeitpunkt t_0,

$$A_{t_0} - \hat{\beta} B_{t_0} \geq \hat{\mu} + th, \tag{23}$$

dann wird ein Handelszyklus begonnen, wobei eine Aktie A verkauft und $\hat{\beta}$ Aktien B gekauft werden. Wird der Handelszyklus zu einem späteren Zeitpunkt $t_1 > t_0$ wieder geschlossen, da für den Spread

$$A_{t_1} - \hat{\beta} B_{t_1} \leq \hat{\mu} \tag{24}$$

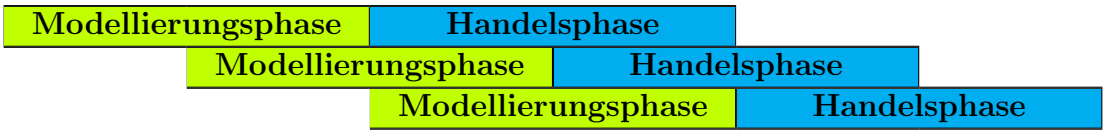
gilt, so lässt sich der erwirtschaftete Gewinn durch

$$\begin{aligned} G &= \hat{\beta}(B_{t_1} - B_{t_0}) + (A_{t_0} - A_{t_1}) \\ &= \hat{\beta}(B_{t_1} - B_{t_0}) + \Big[(\hat{\beta} B_{t_0} + \hat{\mu} + \hat{\epsilon}_{t_0}) - (\hat{\beta} B_{t_1} + \hat{\mu} + \hat{\epsilon}_{t_1})\Big] \\ &= \hat{\epsilon}_{t_0} - \hat{\epsilon}_{t_1} \geq th \end{aligned} \tag{25}$$

berechnen. Der Fehler $\hat{\epsilon}_{t_0}$ muss wegen Ungleichung (23) größer oder gleich den Schwellwert th sein und der Fehler $\hat{\epsilon}_{t_1}$ muss wegen Ungleichung (24) kleiner oder gleich Null sein. Daraus ergibt sich, dass in diesem Fall der minimale Profit durch th gegeben ist. Das selbe Resultat erhält man, falls der Spread zum Zeitpunkt t_0 kleiner als sein Langzeit-Gleichgewicht ist, vgl. Lin et al. (2006).
Bevor wir uns im nächsten Kapitel genauer mit den in der Machbarkeitsstudie durchgeführten Modellierungs- und Auswahlverfahren sowie den verwendeten Handelsstrategien beschäftigen, soll noch die offene Frage geklärt werden, ob Pairs-Trading bei bestehender Kointegrationsbeziehung die Definition einer statistischen Arbitrage-Möglichkeit erfüllt. Chiu und Wong (2012) weisen dies für gewisse zeitkonsistente *Mean-Variance* Strategien nach, wobei sie zusätzlich von einer risikolosen Anleihe mit positivem Return ausgehen.

4 Handelsstrategie

In den letzten beiden Kapiteln wurden einige Pairs-Trading Ansätze untersucht und deren Handelsstrategien vorgestellt. Insbesondere haben wir dabei festgestellt, dass die Mittelwertrückkehr der wichtigste Aspekt ist, denn durch sie ist der erwartete Gewinn positiv. Folglich ist es für Investoren wichtig, sich dieser Eigenschaft - in unserem Falle der Kointegrationsbeziehung - sicher zu sein. Um dies zu gewährleisten werden zunächst die historischen Preise aller zu betrachtenden Aktientitel auf Kointegration untersucht und mittels anderer statistischer Verfahren „ausgedünnt“ , da nicht alle aufgefundenen Aktienpaare mit Kointegrationsbeziehung zum Handeln geeignet sind. Diese zeitlich beschränkte Phase wird als Modellierungsphase bezeichnet. Anschließend werden die zum Handeln akzeptierten Paare ins Portfolio[11] aufgenommen und in der direkt auf die Modellierungsphase folgenden Handelsphase potentiell gehandelt. Diese Phase ist ebenfalls zeitlich beschränkt, um keine Verluste durch eventuell auftretende Änderungen der Kointegrationsbeziehung zu erleiden. Parallel zur Handelsphase läuft bereits überlappend die nächste Modellierungsphase, sodass zu jeder Zeit verfügbare Aktienpaare im Portfolio enthalten sind. Schematisch ist dies in der folgenden Abbildung dargestellt:



4.1 Modellierungsphase

Als Modellierungsphase bezeichnen wir jenen (historischen) Zeitraum $\{1, 2, \ldots, T_1\}$, auf welchem alle verfügbaren Aktienkurse auf bestehende Kointegrationsbeziehungen untersucht werden. Dazu wird eines der in Abschnitt 3.3 vorgestellten Testverfahren auf alle möglichen Aktienpaare angewandt, wobei die Aktienkurse zunächst auf eine Integration von Grad Eins überprüft werden. Bei einer Anzahl von $N \geq 2$ Aktientiteln entspricht dies $N(N-1)$ Tests, falls die Teststatistik abhängig von der gewählten Reihung der Variablen abhängt wie dies beim Engle-Granger Verfahren und bei der Phillips-Ouliaris

[11] Als Portfolio bezeichnen wir beim Pairs-Trading die Menge alle Aktienpaare, die wir zum Handeln akzeptieren. Das sind jene Paare, in deren Kointegrationsbeziehung das meiste Vertrauen gesteckt wird.

$\hat{P}_u$-Statistik der Fall ist. Ist der verwendete Test unabhängig von der Reihenfolge der Variablen, wie beispielsweise der Johansen-Test oder die Phillips-Ouliaris $\hat{P}_z$-Statistik, dann sind $N(N-1)/2$ Tests durchzuführen.
Das Problem an Kointegrationsbeziehungen bei Aktienkursen ist, dass diese oft nur von sehr kurzfristiger Natur sind. Erfüllt beispielsweise ein kointegriertes Aktienpaar A und B in der Modellierungsphase die Regressionsgleichung

$$A_t = \hat{\mu} + \hat{\beta} B_t + \hat{\epsilon}_t,$$

wobei $\hat{\mu}$ und $\hat{\beta}$ die geschätzten Kointegrationsparameter und $\hat{\epsilon}_t$ die geschätzten Residuen sind, so muss diese Beziehung in der unmittelbar folgenden Handelsphase nicht mehr (auf diese Art) existieren. Wir unterscheiden hierbei zwei mögliche Fälle:

(1) Das Aktienpaar A und B erfüllt auch in der Handelsphase eine Kointegrationsbeziehung

$$A_t = \tilde{\mu} + \tilde{\beta} B_t + \tilde{\epsilon}_t,$$

wobei sich mindestens einer der Parameter $\tilde{\mu}$ bzw. $\tilde{\beta}$ signifikant vom bereits geschätzten Parameter $\hat{\mu}$ bzw. $\hat{\beta}$ unterscheidet. Anders ausgedrückt ist der Fehler

$$A_t - (\hat{\mu} + \hat{\beta} B_t)$$

in der Handelsphase nicht stationär, womit auch eine Mittelwertrückkehr nicht mehr zu erwarten ist.

(2) Das Aktienpaar A und B erfüllt in der Handelsphase keine Kointegrationsbeziehung, d.h. jede mögliche Linearkombination der beiden Kurse ist nicht stationär.

Die folgenden Abbildungen zeigen jeweils Beispiele, wie der mittelwertbereinigte Spread $\hat{\epsilon}_t$ in den angesprochenen Fällen verläuft, wobei die vertikale blaue Linie das Ende der Modellierungsphase kennzeichnet.

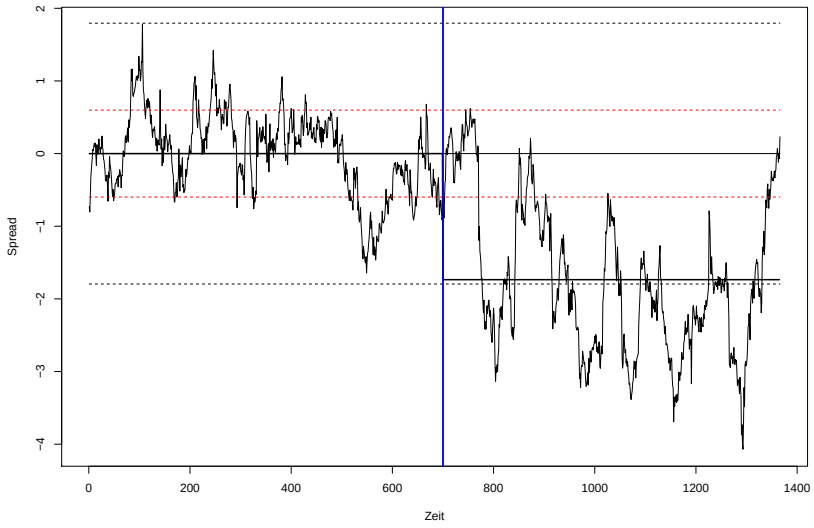


Abbildung 3: Der Mittelwert des Spreads ändert sich deutlich (schwarze Linie).

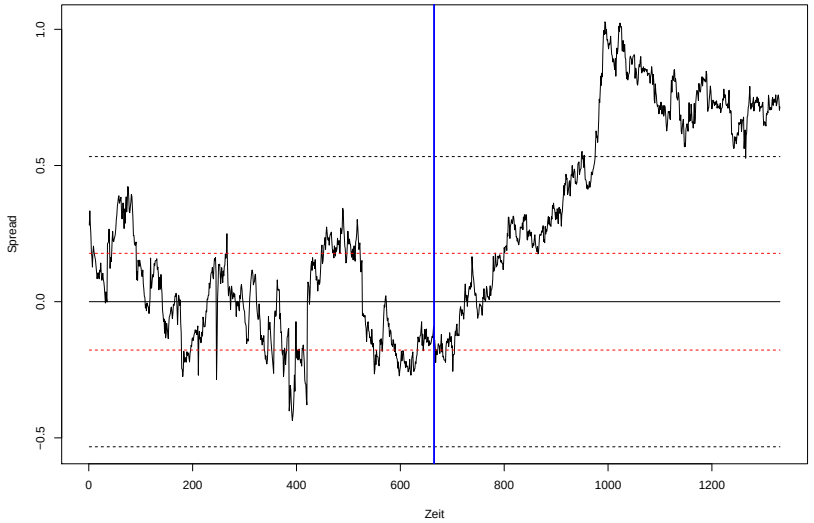


Abbildung 4: Die Kointegrationsbeziehung ist vollständig zusammengebrochen.

Es ist deutlich zu erkennen, dass die zugehörigen Aktienpaare nicht mehr zum Handeln verwendet werden sollten - zumindest nicht mit unveränderten Parametern $\hat{\mu}$ bzw. $\hat{\beta}$. Eine Möglichkeit dieses Problem zu umgehen ist eine adaptive Schätzung von μ bzw. β und wird in Abschnitt 4.2.4 im Zusammenhang mit dem Kalman-Filter besprochen. Aus diesem Grund werden alle es kointegriert klassifizierten Aktienpaare an den Kointegrationstest anschließend durch geeignete Bewertungskriterien auf deren Stabilität

überprüft. Nur jene Paare, deren Kointegrationsbeziehung als stabil genug erscheint, werden ins Portfolio aufgenommen und in der auf die Modellierungsphase folgenden Handelsphase zum Handeln verwendet.

4.1.1 Auswahlkriterien

Um in der Handelsphase keine Verluste zu machen ist es besonders wichtig, dass die Kointegrationsbeziehungen der identifizierten Aktienpaare in der Handelsphase weiterhin stabil bleiben und sich die Kointegrationsparameter μ und β nicht stark ändern. Die Anzahl der gefundenen Paare - d.h. jene Paare, welche bereits aufgrund eines Kointegrationstests als kointegriert eingestuft sind - soll nun durch die Auswertung weiterer Teststatistiken verringert werden, sodass nur die aussichtsreichsten Paare ins Portfolio aufgenommen werden. Die folgende Liste beinhaltet klassische Indikatoren, welche in der Literatur sehr häufig verwendet werden, vgl. etwa Andrade et al. (2005), Caldeira und Moura (2012), Miao (2014) oder Rad et al. (2015).

(1) Der p-Wert des verwendeten Kointegrationstests

(2) Die Distanz der normierten Aktienpreise (vgl. Abschnitt 2.1.1)

$$D_{A,B} := \frac{1}{T_1} \sum_{i=1}^{T_1} \left(\frac{A_t}{A_1} - \frac{B_t}{B_1} \right)^2$$

(3) Der geschätzte Korrelationskoeffizient der Preise

$$\frac{\frac{1}{T_1} \sum_1^{T_1} (A_t - \bar{A})(B_t - \bar{B})}{\sqrt{\frac{1}{T_1} \sum_1^{T_1} (A_t - \bar{A})^2 \frac{1}{T_1} \sum_1^{T_1} (B_t - \bar{B})}},$$

wobei $\bar{A}$ und $\bar{B}$ die jeweiligen Preis-Mittelwerte der Aktien A und B bezeichnen.

(4) Die Autokorrelation der geschätzten Residuen $\hat{\epsilon}_t = A_t - \hat{\beta} B_t - \hat{\mu}$ zum Lag 1

$$\frac{\frac{1}{T_1} \sum_1^{T_1} \hat{\epsilon}_t \hat{\epsilon}_{t-1}}{\frac{1}{T_1} \sum_1^{T_1} \hat{\epsilon}_t^2}$$

(5) Ein Maß für die Ähnlichkeit der beiden CAPM-beta[12], β^A und β^B, beispielsweise $\left|\beta^A - \beta^B\right|$ oder $\left|\frac{\beta^A}{\beta^B} - 1\right|$

(6) Die (annualisierte) in-sample Sharpe-Ratio der beiden Aktien A und B bzw. der in-sample Pairs-Trading Strategie für das Paar A und B. Die Sharpe-Ratio ist definiert durch

$$SR := \frac{\mathbb{E}\left[r^a - r^f\right]}{\sqrt{\mathbb{V}\left[r^p - r^f\right]}},$$

wobei r^p den Return des Portfolios - beispielsweise einer einzelne Aktie - oder einer Pairs-Trading Strategie und r^f den risikolosen Zinssatz angibt (meist wird $r^f = 0$ gesetzt).

Der p-Wert des verwendeten Kointegrationstests ist ein logisches erstes Kriterium für die Auswahl der Aktienpaare. Abhängig von der verwendeten Anzahl unterschiedlicher Aktientitel kann jedoch eine Vorauswahl potentiell kointegrierter Paare notwendig sein, um die Berechnungszeiten zu verkürzen. Ein schnelles Auswahlverfahren ist - bedingt durch sehr kurze Modellierungsphasen - vor allem beim sogenannten Hochfrequenzhandel (auf englisch *High-Frequency Trading*) essenziell. Miao (2014) schlägt hierfür die Verwendung des geschätzten Korrelationskoeffizienten bzw. der Distanz vor, wobei in seiner Arbeit nur Aktienpaare auf Kointegration untersucht werden, deren Korrelationskoeffizient größer oder gleich 0.90 ist. Anschließend werden abhängig vom p-Wert es Kointegrationstests zehn Paare ins Portfolio aufgenommen und zum Handeln verwendet. Eine Motivation für die Betrachtung von Paaren mit ähnlichen CAPM-betas liefert beispielsweise Kishore (2012) dadurch, dass von Aktientiteln mit ähnlichem CAPM-beta eine ähnliche Beeinflussung durch den Markt erwartet wird. Weiters zeigt das folgende Beispiel, dass sich durch die Wahl einer geeigneten Hedge-Ratio eine marktneutrale Pairs-Trading Strategie konstruieren lässt.

Beispiel 5 *Betrachten wir zwei - nicht notwendigerweise kointegrierte - Aktientitel A und B mit CAPM-betas β_A bzw. β_B ungleich Null. Die Returns der Aktien erfüllen somit*

[12]Die Abkürzung CAPM steht für *Capital Asset Pricing Model* - ein Preismodell für Kapitalgüter. Für eine Aktie A berechnet sich das CAPM-β_A aus der linearen Regression $r_t^A = \alpha + \beta_A r_t^M + \mu_t$, wobei $r_t^A = (A_t - A_{t-1})/A_{t-1}$ die Returns von A und r_t^M die Returns des Marktes bezeichnet.

die Regressionsgleichungen

$$r_t^A = \alpha_A + \beta_A r_t^M + \mu_t$$
$$r_t^B = \alpha_B + \beta_B r_t^M + \nu_t,$$

wobei r_t^M die Returns des Marktes und μ_t bzw. ν_t die Residuen dieser beiden Regressionen sind. Die Returns r_t^P einer (buy-and-hold) Pairs-Trading Strategie gegeben durch

$$1 \text{ Long} - \text{Position in A und } \gamma \text{ Short} - \text{Positionen in B}$$

erfüllen

$$r_t^P = r_t^A - \gamma r_t^B = (\alpha_A + \alpha_B) + (\beta_A - \gamma\beta_B) r_t^M + \mu_t - \gamma\nu_t.$$

Das CAPM-beta dieser einfachen Strategie ist somit durch $\beta_P := \beta_A - \gamma\beta_B$ gegeben, welches durch die Wahl

$$\gamma = \frac{\beta_A}{\beta_B}$$

gleich Null wird und somit eine marktneutrale Strategie liefert.

Caldeira und Moura (2012) verwenden jene 20 Aktienpaare, deren Pairs-Trading Strategie die höchsten annualisierten in-sample Sharpe-Ratios aufweisen und welche zunächst sowohl mittels Engle-Granger Verfahrens als auch mittels Johansen-Tests als kointegriert klassifiziert wurden. Auch in ihrer Arbeit werden möglichst marktneutrale Strategien verwendet werden. Die hohe in-sample Sharpe-Ratio der Handelsstrategien deutet auf ein gutes Verhältnis der Returns zum eingegangenen Risiko hin.
Zwei im Praxisteil dieser Arbeit implementierte Stabilitätstests sind folgendermaßen gegeben:

Stabilitätstests Der erste Stabilitätstest untersucht die Kointegrationsbeziehung eines auf der gesamten Modellierungsphase $\{1, 2, \ldots, T_1\}$ als kointegriert klassifiziertes Aktienpaares A und B auf den beiden „Hälften“

$$\{1, 2, \ldots, \lfloor T_1/2 \rfloor\} \quad \text{bzw.} \quad \{\lfloor T_1/2 \rfloor + 1, \ldots, T_1\}.$$

Nur jene Paare, die auf beiden Teilfenstern zumindest zum 10 %-Niveau als kointegriert gelten werden potentiell ins Portfolio aufgenommen.

Der zweite Stabilitätstest betrachtet ein rollendes Zeitfenster der Länge $T_R < T_1$, welches innerhalb der Modellierungsphase überlappend verschoben wird. Gibt $1 \leq L \leq T_1 - T_R$ die (konstante) Verschiebung an, dann werden auf allen Fenstern

$$\{1 + kL, \ldots, T_R + kL\}, \quad k \in \{0, \ldots, \lfloor (T_1 - T_R)/L \rfloor\}$$

die Regressionsparameter μ und β geschätzt. Nur jene Paare mit geringer Variation dieser Schätzungen werden potentiell ins Portfolio aufgenommen.

4.2 Handelsphase

Die Handelsphase bezeichnet jenes diskrete Zeitfenster $\{T_1+1, \ldots, T\}$, welches unmittelbar an die Modellierungsphase anschließt. Alle sich im Portfolio befindenden Aktienpaare wurden in der Modellierungsphase als kointegriert klassifiziert und werden unabhängig voneinander gehandelt. Für ein solches Aktienpaar A und B ist der Spread

$$A_t - \hat{\beta} B_t = \hat{\mu} + \hat{\epsilon}_t$$

schwach-stationär mit Mittelwert $\hat{\mu}$. Verwendet wird nun die Handelsstrategie aus Abschnitt 3.4 derart, dass immer nur ein Handelszyklus gleichzeitig stattfinden darf. Das bedeutet, dass ein neuer Handelszyklus erst begonnen werden darf, wenn ein bestehender zuvor glattgestellt wurde. Ist dies der Fall, dann werden Positionen folgendermaßen eingegangen

$$\begin{aligned} A_t - \hat{\beta} B_t \leq \hat{\mu} - th \qquad & \Rightarrow 1 \text{ Short-Position in A, } \gamma \text{ Long-Positionen in B} \\ A_t - \hat{\beta} B_t \geq \hat{\mu} + th \qquad & \Rightarrow 1 \text{ Long-Position in A, } \gamma \text{ Short-Positionen in B} \end{aligned}$$

wobei γ die Hedge-Ratio bezeichnet und $th > 0$ ein positiver Schwellwert ist. Sobald der Spread erneut seinen Mittelwert $\hat{\mu}$ durchschreitet oder das Ende der Handelsphase T erreicht ist, werden alle offenen Positionen glattgestellt. Diese Handelsstrategie wird vor allem durch den Wert th gesteuert. Ist der Wert th zu groß, dann kommt es seltener oder eventuell nie zu einem Handelszyklus. Ist er zu klein, dann werden durch - sehr wahrscheinlich existierende - Transaktionskosten Verluste gemacht. Daher ist die Wahl eines geeigneten Schwellwertes essenziell für die Handelsstrategie und wird im nächsten Abschnitt behandelt.

4.2.1 Der Schwellwert „th“

Der Schwellwert $th > 0$ ist üblicherweise ein Vielfaches der Standardabweichung des historischen Spreads,

$$\hat{\sigma}_{sp} = \sqrt{\frac{1}{T_1} \sum_{t=1}^{T_1} \hat{\epsilon}_t},$$

$th = j\hat{\sigma}_{sp}$ mit $j > 0$. Es sind jedoch auch nicht konstante Werte möglich, beispielsweise durch eine Neuberechnung der Standardabweichung des Spreads bis zum aktuellen Zeitpunkt t_0

$$\hat{\sigma}_{sp,t_0} = \sqrt{\frac{1}{T_1} \sum_{t=0}^{T_1-1} \hat{\epsilon}_{t_0-t}},$$

wobei solche „Updates“ üblicherweise nach einer gewissen Anzahl von Handelszeitpunkten oder nach Abschluss eines Handelszyklus erfolgen.
Wichtig ist auf jeden Fall, dass der Schwellenwert mindestens so groß gewählt wird, dass trotz anfallender Transaktionskosten bei jedem vollständigen Handelszyklus Profit gemacht wird. Hierfür ist eine Abschätzung der Transaktionskosten zum Zeitpunkt der Glattstellung notwendig, welche beispielsweise über die durchschnittlichen (historischen) Preise der beiden Aktientitel getroffen werden kann.

Transaktionskosten Da Pairs-Trading Strategien auf häufigem Handeln basieren, spielen die Transaktionskosten eine wichtige Rolle. Sie liefern eine untere Schranke für den Schwellwert th, da der erwartete Gewinn die Transaktionskosten auf jeden Fall übersteigen sollte. Unterschieden wird hier zwischen Fixkosten und variablen Kosten. Bei den Fixkosten handelt es sich um einen gewissen Betrag, der unabhängig vom gehandelten Volumen bei jeder Transaktion zu entrichten sind. Die variablen Kosten hingegen sind ein gewisser Prozentsatz des Wertes der eingegangenen Positionen.

4.2.2 Die Hedge-Ratio γ

Die Hedge-Ratio γ wird beim Kointegrations-Ansatz üblicherweise dem geschätzten Parameter $\hat{\beta}$ gleichgesetzt, $\gamma = \hat{\beta}$. In der Abschätzung (25) aus Abschnitt 3.4 haben wir bereits gesehen, dass dadurch für jeden Handelszyklus der minimale Gewinn „th“ beträgt. Es sind jedoch auch andere Werte möglich, die je nach Vorlieben des Investors

gewählt werden können. Durch eine geeignete Wahl von γ lässt sich beispielsweise eine marktneutrale Handelsstrategie konstruieren (vgl. Beispiel 5). Vor allem in theoretischen Abhandlungen werden häufig Pairs-Trading Strategien betrachtet, bei denen die Einnahmen durch den Leerverkauf der einen Aktie zum Kaufen der anderen Aktie verwendet werden. Dadurch ist zumindest theoretisch kein Eigenkapital notwendig. In der Praxis funktioniert dies jedoch nicht so einfach, da sowohl Transaktionskosten als auch meist notwendige Sicherheitsleistungen für Leerverkäufe berücksichtigt werden müssen.
Werden für die Untersuchung auf existierende Kointegrationsbeziehungen die logarithmierten Kurswerte der Aktien verwendet - d.h. es wird eine Regression

$$\log A_t = \tilde{\mu} + \tilde{\beta} \log B_t + \tilde{\epsilon}_t \tag{26}$$

betrachtet - dann wird üblicherweise nicht der Koeffizient $\tilde{\beta}$ als Hedge-Ratio verwendet. Motivieren lässt sich dies folgendermaßen: Bei einer Regression der Preise,

$$A_t = \hat{\mu} + \hat{\beta} B_t + \epsilon_t,$$

entspricht $\hat{\beta}$ der formalen partiellen Ableitung von A_t bzgl. B_t,

$$\frac{\partial A_t}{\partial B_t} = \hat{\beta},$$

und somit der momentanen Änderungsrate von A_t bzgl. B_t. Dies erkennt man auch einfach daran, dass $\hat{\beta}$ die Steigung der Regressionsgerade ist. Wird jedoch die Regression der logarithmierten Preise betrachtet, dann ergibt sich für diese formale Ableitung unter Verwendung von Gleichung (26)

$$\begin{aligned}
\frac{\partial A_t}{\partial B_t} &= \frac{\partial \exp\left[\tilde{\mu} + \tilde{\beta} \log B_t + \tilde{\epsilon}_t\right]}{\partial B_t} \\
&= \tilde{\beta} \frac{1}{B_t} \exp\left[\tilde{\mu} + \tilde{\beta} \log B_t + \tilde{\epsilon}_t\right] \\
&= \tilde{\beta} \frac{A_t}{B_t}.
\end{aligned}$$

Der Parameter $\tilde{\beta}$ entspricht somit einer Schätzung der Elastizität der Preise A_t bzgl. B_t,

$$E_{A_t,B_t} = \frac{\partial A_t}{\partial B_t} \cdot \frac{B_t}{A_t}.$$

Die Hedge-Ratio $\gamma := \tilde{\beta} A_t / B_t$ ist somit nicht konstant (wie dies bei Betrachtung der tatsächlichen Preise der Fall ist), sondern sie hängt von den Preisen der beiden Aktientitel zum Zeitpunkt t ab. Wird zu einem Zeitpunkt t_0 ein Handelszyklus geöffnet, so wird jedoch die Hedge-Ratio bis zum Ende dieses Handelszyklus konstant gehalten, d.h. offene Positionen werden mit der selben Hedge-Ratio glattgestellt.

4.2.3 Sicherheitsmechanismen

Üblicherweise werden bei Pairs-Trading Strategien unterschiedliche Sicherheitsmechanismen implementiert, um vor allem eine Veränderung der Kointegrationsbeziehung rechtzeitig zu erkennen. Die beiden in Abschnitt 4.1 erwähnten Fälle - d.h. eine signifikante Änderung der Parameter μ und/oder β bzw. ein vollständiger Zusammenbruch der Kointegrationsbeziehung - lassen sich nicht vorhersagen. Oft reichen Gerüchte - beispielsweise eines Übernahmeangebotes - aus, um die Aktienkurse stark zu beeinflussen und eine bestehende Kointegrationsbeziehung zu verändern. Beim Pairs-Trading wird dies meist durch sogenannte *Stop-Loss* Schranken erreicht, welche folgendermaßen gegeben sind:

(1) Ein Schwellwert $\bar{th} > th$ wird eingeführt. Gilt für den Spread

$$\left| A_t - \hat{\beta} B_t \right| \geq \bar{th}$$

dann werden eventuell eingegangene Positionen sofort glattgestellt, obwohl hierbei wegen $\bar{th} > th$ auf jeden Fall Verlust gemacht wird. Das Aktienpaar A und B wird außerdem für die aktuelle Handelsphase aus dem Portfolio entfernt, d.h. dieses Paar kann in dieser Phase nicht mehr gehandelt werden.

(2) Eine maximale Zeit $t_{max} > 0$ wird festgelegt. Wird zu einem Zeitpunkt t_0 ein Handelszyklus begonnen, dann wird spätestens zum Zeitpunkt $t_1 = t_0 + t_{max}$ der Handelszyklus beendet. Dies tritt dann ein, falls in diesem Zeitraum keine Mittelwertrückkehr eingetreten ist und die oben erwähnte Schranke $\bar{th}$ nicht durchschritten wurde. Auch in diesem Fall wird das Aktienpaar A und B für die aktuelle Handelsphase aus dem Portfolio entfernt.

Entfernt sich der Spread zu weit vom seinem Gleichgewicht $\hat{\mu}$ oder kehrt er zu lange nicht zu diesem zurück, dann deutet dies auf eine veränderte oder zusammengebrochene Kointegrationsbeziehung hin. Ein weiteres Handeln mit diesem Paar wäre ohne neu durchgeführten Kointegrationstest und neu berechnete Parameter $\hat{\mu}$ und $\hat{\beta}$ zu riskant, folglich wird dieses Aktienpaar für diese Handelsphase aus dem Portfolio entfernt. Wird

es in einer späteren Modellierungsphase erneut als kointegriert klassifiziert, so wird es auch erneut ins Portfolio für die darauf folgende Handelsphase aufgenommen.

4.2.4 Adaptive Schätzung mittels Kalman-Filter

Wie bereits in den vorangegangene Abschnitten erwähnt, ist eine stabile Kointegrationsbeziehung beim Pairs-Trading äußerst wichtig, kann jedoch nicht garantiert werden. Das Kalman-Filter ist ein möglicher Ansatz, dieses Problem in den Griff zu bekommen. Hierbei werden die Regressions-Parameter β und μ nicht als konstant aufgefasst, sondern als Zeit-dynamische Variablen β_t bzw. μ_t. Betrachtet wird hierfür das folgende Zustandsraum-Modell

$$\mu_{t+1} = \mu_t + \eta_t \tag{27}$$

$$\beta_{t+1} = \beta_t + \nu_t \tag{28}$$

$$A_t = \mu_t + \beta_t B_t + \epsilon_t, \tag{29}$$

wobei $\xi_t := (\eta_t, \nu_t, \epsilon_t)'$ ein weißes Rauschen mit Kovarianzmatrix

$$\mathbb{V}\left[\xi_t \xi_t'\right] = \begin{pmatrix} \sigma_\eta^2 & 0 & 0 \\ 0 & \sigma_\nu^2 & 0 \\ 0 & 0 & \sigma_\epsilon^2 \end{pmatrix} \tag{30}$$

ist. Die beiden Random-Walk Gleichungen (27) und (28) beschreiben die Zustandsgrößen β_t bzw. μ_t dieses Modells und werden Zustandsgleichungen genannt. Die beobachtete Variable ist A_t und die zugehörige Gleichung (29) wird als Beobachtungsgleichung bezeichnet.
Die sogenannte Tuning-Parameter des Kalman-Filter, d.h. die Varianzen $(\sigma_\eta^2, \sigma_\nu^2, \sigma_\epsilon^2)$, werden mittels Maximum-Likelihood Verfahrens in der Modellierungsphase bestimmt. Die Startwerte für $\hat{\mu}_t$ und $\hat{\beta}_t$ stammen von der geschätzten Kointegrationsbeziehung des Aktienpaares. Bezeichnen wir mit

$$\mathcal{Y}_k := \sigma\{A_1, \ldots, A_k, B_1, \ldots, B_k\}$$

die von den beobachteten Preisen bis zum Zeitpunkt k aufgespannte Sigma-Algebra, dann berechnet das Kalman-Filter rekursiv die best-mögliche Schätzungen der Zustands-

Parameter (μ_t, β_t) gegeben dieser Preise,

$$(\hat{\mu}_t, \hat{\beta}_t) = \mathbb{E}\left[(\mu_t, \beta_t)|\mathcal{Y}_t\right].$$

Zusätzlich zu diesen Schätzungen erhält man auch eine Schätzung für den Spread,

$$A_t - \hat{\beta}_t B_t = \hat{\mu}_t + \hat{\epsilon}_t, \tag{31}$$

welchen wir fortan als *adaptiv geschätzten Spread* bezeichnen werden.

Handelsstrategie Die Handelsstrategie funktioniert analog zur üblichen Pairs-Trading Strategie beim Kointegrations-Ansatz. Die Signale, wann ein Handelszyklus gestartet wird, werden durch den adaptiv geschätzten Spread

$$A_t - \hat{\beta}_t B_t = \hat{\mu}_t + \hat{\epsilon}_t$$

und einen positiven Schwellerwert $th > 0$ gesteuert, der üblicherweise als ein Vielfaches die Standardabweichung

$$\hat{\sigma} = \sqrt{\frac{1}{T_1}\sum_{t=1}^{T_1} \hat{\epsilon}_t^2}$$

des adaptiven Spreads definiert ist, d.h. $th := \delta\hat{\sigma}$ mit $\delta > 0$. Ist der adaptiv geschätzte Spread zu einem Zeitpunkt $t_0 \in \{T_1 + 1, \ldots, T\}$ hinreichend weit von seinem adaptiv geschätzten Mittelwert $\hat{\mu}_t$ entfernt, dann wird ein Handelszyklus folgendermaßen begonnen

$$\begin{aligned} A_{t_0} - \hat{\beta}_{t_0} B_{t_0} \leq \hat{\mu}_{t_0} - th \qquad &\Rightarrow 1 \text{ Short-Position in A, } \gamma \text{ Long-Positionen in B} \\ A_{t_0} - \hat{\beta}_{t_0} B_{t_0} \geq \hat{\mu}_{t_0} + th \qquad &\Rightarrow 1 \text{ Long-Position in A, } \gamma \text{ Short-Positionen in B} \end{aligned}$$

wobei die Hedge-Ratio γ normalerweise gleich $\hat{\beta}_{t_0}$ gesetzt wird[13]. Ab diesem Zeitpunkt werden $\hat{\mu}_t$ und $\hat{\beta}_t$ im adaptiven Spread konstant $\hat{\mu}_{t_0}$ bzw. $\hat{\beta}_{t_0}$ gesetzt. Ist beispielsweise der erste Fall eingetrete, d.h. $A_{t_0} - \hat{\beta}_{t_0} B_{t_0} \leq \hat{\mu}_{t_0} - th$, so erfolgt eine Glattstellung der

[13] Werden die logarithmierten Preise fürs Erstellen des adaptiven Spreads verwendet, dann sollte die Hedge-Ratio wie in Abschnitt 4.2.2 motiviert gleich $\hat{\beta}_{t_0} A_{t_0}/B_{t_0}$ gesetzt werden.

Positionen, falls zu einem Zeitpunkt $t_1 > t_0$

$$A_{t_1} - \hat{\beta}_{t_0} B_{t_1} \geq \hat{\mu}_{t_0}$$

gilt. Der zweite Fall verläuft analog dazu.
Wichtig zu beachten ist, dass bei Verwendung von Stop-Loss Schranken wie in Abschnitt 4.2.3 nicht *nur* $t\bar{h} > th$ verlangt werden muss. Der Schwellwert th gibt den Mindestabstand an, den der adaptiv geschätzte Spread von seinem geschätzten Mittelwert $\hat{\mu}_t$ haben muss, damit ein Handelszyklus begonnen wird. Da die Standardabweichung des adaptiv geschätzten Spreads jedoch im Vergleich zu einem *normalen* Spread relativ klein ist, muss $t\bar{h}$ wesentlich größer gewählt werden. Üblicherweise wird hierfür die Standardabweichung des (historischen) *gehandelten* Spreads herangezogen, welche durch

$$\bar{\sigma} := \sqrt{\frac{1}{T_1} \sum_{t=1}^{T_1} (A_{t_0 - t} - \hat{\beta}_{t_0} B_{t_0 - t})^2}$$

gegeben ist. Die Stop-Loss Grenze $t\bar{h}$ kann nun als ein Vielfaches dieser Standardabweichung definiert werden, $t\bar{h} := \bar{\delta}\bar{\sigma}$ mit $\bar{\delta} > 0$. Der Verlauf eines solchen (mittelwertbereinigten) adaptiv geschätzten Spreads inklusive der soeben besprochenen Handelsstrategie ist in der folgenden Abbildung zu sehen, wobei das grüne Band die historische Standardabweichung - berechnet aus den vorangegangenen 4 Wochen - und das rote Band die Stop-Loss Schranke mit $t\bar{h} := 3\bar{\sigma}$ sind. Die Handelszyklen sind jeweils grau hinterlegt.

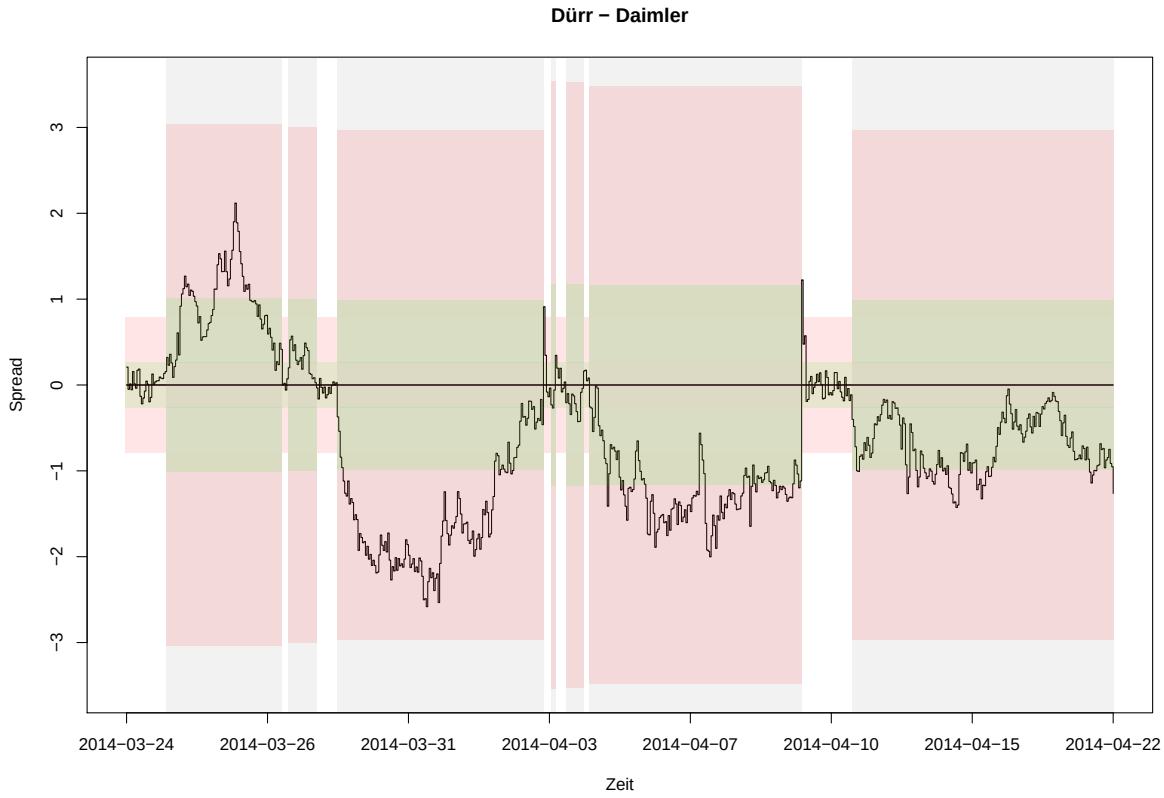


Abbildung 5: Der adaptiv geschätzte Spread in der Handelsphase.

Man erkennt deutlich die größere Standardabweichung während der Handelszyklen und die sich daraus ergebenden größeren Schwellwerte th und $\bar{th}$. Die nächste Grafik zeigt rot-strichliert die adaptiv geschätzten Parameter $\hat{\beta}_t$ und $\hat{\mu}_t$. In schwarz sind jene Werte zu erkennen, welche für die Erzeugung des Spreads aus Abbildung 5 verwendet wurden, d.h. die in Handelszyklen konstant gehaltenen Parameter $\hat{\beta}_t$ und $\hat{\mu}_t$.

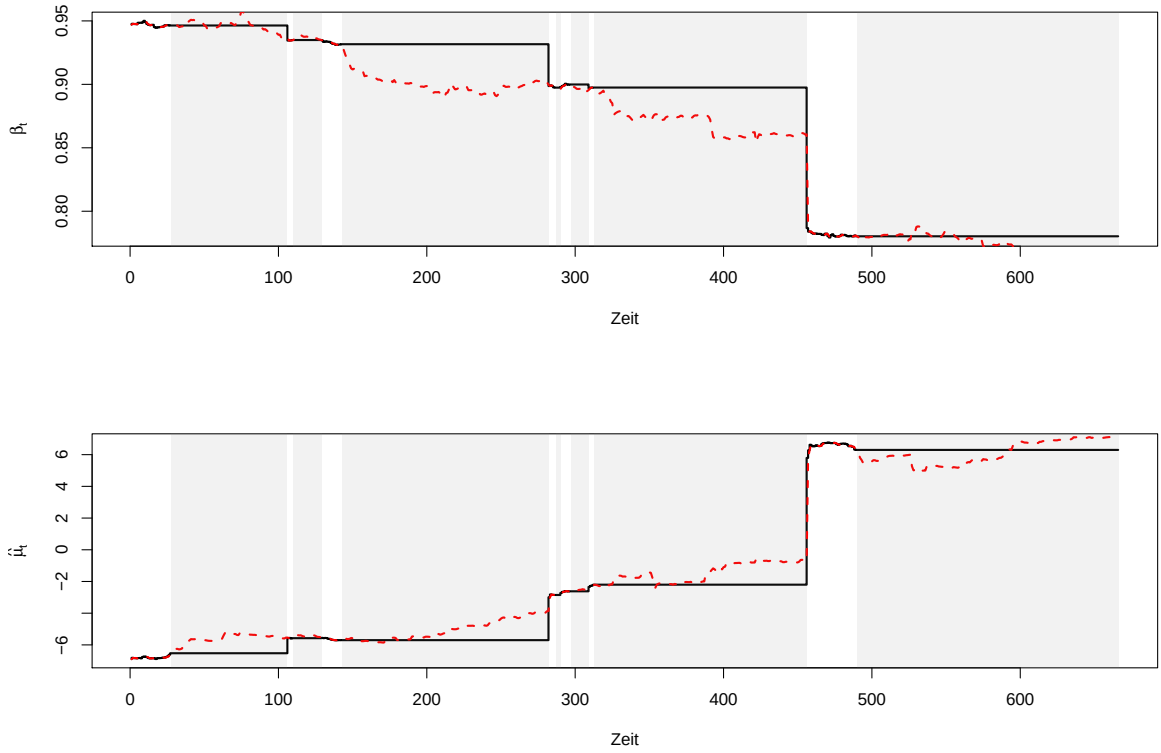


Abbildung 6: Die adaptiv geschätzten Parameter $\hat{\beta}_t$ und $\hat{\mu}_t$.

5 Praxisteil

In diesem Abschnitt sollen die zuvor diskutierten Pairs-Trading Strategien basierend auf dem Kointegrations-Ansatz und anhand historischer Aktienkurse getestet werden. Alle Untersuchungen erfolgen mit der freien Programmiersprache *R* in seiner jeweilig aktuellen Version. Zunächst folgt eine kurze Beschreibung der verwendeten Daten. Insbesondere wird untersucht, ob die Kurse integriert erster Ordnung sind und ob sich Kointegrationsbeziehungen finden lassen. Anschließend werden die genauen Parameter für die in Abschnitt 4 besprochenen Handelsstrategien angegeben, mit welchen diese Strategien getestet wurden. Am Ende dieses Abschnittes befindet sich eine Zusammenfassung der erhaltenen Resultate.

5.1 Datenbeschreibung

Für die dieser Diplomarbeit zugrundeliegende Machbarkeitsstudie wurden Aktientitel des deutschen Aktienindexes *DAX* und des sogenannten *Mid-Cap-DAX* (kurz *MDAX*) verwendet, welche uns von der Firma PS Quant OG (i.G.) zur Verfügung gestellt wurden. Beim DAX handelt es sich um die 30 umsatzstärksten Unternehmen in Deutschland - gelistet an der Frankfurter Wertpapierbörse - und er gilt als Leitindex für den deutschen Aktienmarkt. Der MDAX umfasst jene 50 deutsche Aktientitel, die bezogen auf ihre Marktkapitalisierung und ihren Börsenumsatz den Titeln des DAX folgen.
Insgesamt wurden die Kursdaten von 61 Unternehmen des DAX bzw. MDAX auf Integration und Kointegration untersucht, wobei die Preise in einem 15-Minuten Intervall von 09:00:00 Uhr bis 17:30:00 Uhr gegeben sind, was 35 Beobachtungen pro Tag entspricht. Die Aktienkurse stehen uns auf dem Zeitraum vom 01.01.2012 bis zum 01.10.2015 zur Verfügung, damit handelt es sich um ca. 33200 Beobachtungen jedes Aktientitels. In Abbildung 7 sind die Kurse des DAX bzw. MDAX abgebildet. Bedingt durch die höhere Anzahl inkludierter Aktientitel liegt der MDAX-Kurs über jenem des DAX. Außerdem finden sich beim MDAX Branchen wieder, welche im DAX nicht vertreten sind, beispielsweise Immobilienfirmen oder Medienkonzerne[14].

[14] Im betrachteten Zeitraum bis Oktober 2015 befanden sich keine Medienkonzerne im DAX. Seit 21.03.2016 ist jedoch die *ProSiebenSat.1 Media SE* Teil des DAX.

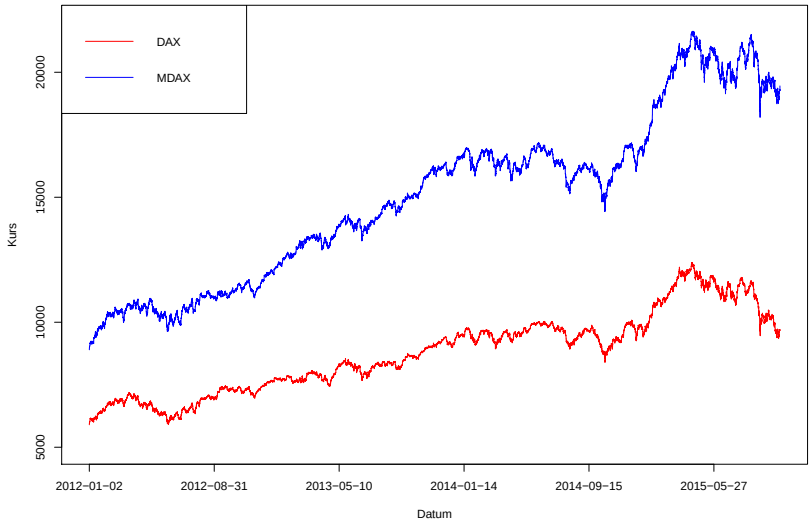


Abbildung 7: Die DAX und MDAX Kurse ab Anfang 2012.

Jede dieser Aktien beinhaltet die folgenden Werte

Time	Close	Volume	High	Low	Open	Date
09:00:00	74.25	58789	74.56	73.58	73.58	02.01.2012

wobei es sich hier um ein Beispiel der *Allianz* handelt. Für die hier besprochenen Untersuchungen auf Kointegration werden die Eröffnungs-Preise (die *Open*-Preise) verwendet. Eine Analyse der anderen Preise - oder beispielsweise der Durchschnittspreise $(High + Low)/2$ - liefert qualitativ die selben Ergebnisse.
Die Kursdaten sind für den gegeben Zeitraum nicht vollständig, da Preisdaten nur dann vorliegen, falls für das jeweilige Unternehmen in dieser Viertelstunde ein Handel stattgefunden hat. Insgesamt ergibt sich die folgende Statistik für die fehlenden Preisdaten pro Aktientitel

Minimum	1. Quartil	Median	Mittelwert	3. Quartil	Maximum
10.00	11.00	20.00	82.33	70.00	947.00

wobei vor allem die Aktientitel *Fielmann* (947 fehlende Werte), *ElringKlinger* (636 fehlende Werte), *Celesio* (580 fehlende Werte) und *Deutsch EuroShop* (402 fehlende Werte) auffallen. Für die folgenden Untersuchungen auf Integration und Kointegration werden die fehlenden Werte linear interpoliert. Eine weitere Möglichkeit ist durch das Beibehalten des letzten Kurswertes gegeben, wobei hier keine qualitativen Veränderungen der Resultate bemerkbar sind.

5.1.1 Existieren Kointegrationsbeziehungen?

In diesem Abschnitt sollen die vorliegenden DAX und MDAX Aktientitel auf mögliche Kointegrationsbeziehungen untersucht werden. Da eine wesentliche Voraussetzung für die Existenz solcher Kointegrationsbeziehungen die Integration der Aktienkurse ersten Grades ist, werden die Kurse zunächst daraufhin überprüft.

Sind die Aktien Integriert? Für diese Untersuchung betrachten wir Zeitfenster mit einer Länge von vier Wochen (entsprechend 20 Handelstagen mit insgesamt ca. 700 Beobachtungen), welche jeweils um vier Wochen verschoben werden. Dadurch erhalten wir insgesamt 48 Zeitfenster, auf welchen die 61 Aktienkurse auf Integration ersten Grades mittels des augmentierten Dickey-Fuller Testes untersucht werden, vgl. Dickey und Fuller (1981). Da wir für die Handelsstrategien auch die logarithmierten Preise betrachten, wird diese Untersuchung ebenfalls für die Log-Preise durchgeführt. Die Ablehnungsrate der Nullhypothese

$$H_0 : \textit{ die Preise/Log} - \textit{Preise sind Integriert vom Grad Eins}$$

ist in der folgenden Tabelle dokumentiert,

Signifikanzniveau	1 %	5 %	10 %
Preise	1.5 %	4.3 %	7.9 %
Log-Preise	1.6 %	4.3 %	8.0 %

wobei die Werte dem prozentualen Anteil der insgesamt $48 \cdot 61 = 2928$ durchgeführten Tests entspricht. Besonders fällt bei dieser Untersuchung die Aktie *Celesio* auf, bei welcher die Nullhypothese „Integriert vom Grad Eins“ zum Signifikanzniveau 10 % auf 11 der 48 Fenstern verworfen wird, was ca. 22 % entspricht.
Man erkennt deutlich, dass die Hypothese „Die (Log-) Preise sind Integriert erster Ordnung“ nur selten abgelehnt wird, die Ablehnungsrate entspricht ungefähr dem Signifikanzniveau des durchgeführten Tests.

Sind die Kursdaten kointegriert? Da die Aktientitel erwartungsgemäß meist als Integriert der Ordnung Eins klassifiziert werden, ist der nächste Schritt die Überprüfung auf existierende Kointegrationsbeziehungen. Hierfür werden alle $61 \cdot 60 = 3660$ möglichen Aktienpaare auf allen 48 Zeitfenstern mittels des Zwei-Schritt-Verfahren und des Johansen-Tests auf Kointegration untersucht. Da das Engle-Granger Verfahren abhängig von der

Reihenfolge des Aktienpaares ist und im Allgemeinen unterschiedliche Ergebnisse liefert, je nachdem welcher Aktientitel als abhängige und welche als unabhängige gewählt wird, führt dies zu $48 \cdot 3660 = 175680$ durchgeführten Tests. Der Johansen-Test hingegen ist unabhängig von der gewählten Reihung des Aktienpaares, d.h. hier wurden $48 \cdot 1830 = 87840$ Tests auf Kointegration durchgeführt.
Die Nullhypothese beider Tests lautet

$$H_0 : \textit{ das Aktienpaar ist nicht Kointegriert}$$

Die Ablehnungsrate dieser Nullhypothese zu den unterschiedlichen Signifikanzniveaus für die Preise und für die logarithmierten Preise ist in der folgenden Tabelle dokumentiert:

Signifikanzniveau	1 %	5 %	10 %
Engle-Granger: Preise	1.6 %	6.4 %	11.2 %
Engle-Granger: Log-Preise	1.6 %	6.5 %	11.4 %
Johansen: Preise	1.6 %	5.7 %	10.3 %
Johansen: Log-Preise	1.7 %	5.7 %	10.4 %

Hier ist deutlich erkennbar, dass die Prozentsätze gefundener kointegrierter Paare nur unwesentlich größer als die jeweiligen Signifikanzniveaus der verwendeten Tests sind. Dies deutet bereits darauf hin, dass bei vielen Paaren die Nullhypothese keiner vorliegenden Kointegrationsbeziehung fälschlicherweise verworfen wurde (Fehler erster Art). Im nächsten Abschnitt wird diese Vermutung durch weitere Tests auf Stabilität der gefundenen Kointegrationsbeziehungen bestätigt.

Bestätigung durch Stabilitätstests Für Pairs-Trading Strategien basierend auf dem Kointegrations-Ansatz ist es nicht nur notwendig, dass Kointegrationsbeziehungen existieren, sondern diese müssen auch für eine gewisse Zeit stabil bleiben. Genauer gesagt ist es essenziell, dass sich die Parameter μ und β der Kointegrationsbeziehung

$$A_t = \mu_t + \beta_t + B_t + \epsilon_t$$

in der Handelsphase nicht stark von jenen aus der Modellierungsphase unterscheiden. Wir testen daher zunächst, wie oft als kointegriert klassifizierte Aktienpaare überhaupt auf benachbarten Fenstern als kointegriert klassifiziert werden. Das Engle-Granger Verfahren und der Johansen-Test liefern dafür zum 10 % Signifikanzniveau die in der folgenden Tabelle dokumentierten Ergebnisse, wobei in der Spalte „#Paare“ die Anzahl

jener Paare angegeben ist, die auf mindestens einem der 48 Zeitfenster als kointegriert klassifiziert wurden. In der Spalte „1 Fenster“ ist jener Prozentsatz dieser Paaren angegeben, welche auf keinen aufeinanderfolgenden Zeitfenstern als kointegriert klassifiziert wurden. In der Spalte „2 Fenster“ sind es jene Paare, welche auf exakt zwei aufeinanderfolgenden Zeitfenstern positiv auf Kointegration getestet wurden. Die restlichen Spalten „3 Fenster“ und „≥4 Fenster“ sind analog zu interpretieren.

	#Paare	1 Fenster	2 Fenster	3 Fenster	≥4 Fenster
Engle-Granger: Preise	3625	62.3 %	31.6 %	5.8 %	0.4 %
Engle-Granger: Log-Preise	3626	61.7 %	32.0 %	5.8 %	0.5 %
Johansen: Preise	1818	58.3 %	34.8 %	6.2 %	0.8 %
Johansen: Log-Preise	1821	56.5 %	35.9 %	6.8 %	0.9 %

Es existieren also Aktienpaare, welche auf mindestens zwei aufeinanderfolgenden Zeitfenstern als kointegriert klassifiziert werden. Die Frage die sich nun stellt ist jedoch, wie stark und wie schnell sich die Koeffizienten μ und β verändern. Zu diesem Zweck werden die Zeitfenster in zwei Hälften geteilt. Auf der ersten Hälfte wird das Regressionsmodell

$$A_t = \hat{\mu} + \hat{\beta} B_t + \hat{\epsilon}_t$$

geschätzt, wobei nur jene Paare betrachtet werden, welche auf dem gesamten Zeitfenster mindestens zum 10 % Signifikanzniveau als kointegriert klassifiziert wurden. Der Spread $A_t - \hat{\beta} B_t$ wird anschließend auf der zweiten Hälfte des Fensters mittels augmentierten Dickey-Fuller Tests auf Stationarität überprüft. Die Ablehnungsraten zu den unterschiedlichen Signifikanzniveaus der Nullhypothese

$$H_0 : \textit{ der Spread ist integriert}$$

ist in der folgenden Tabelle ersichtlich:

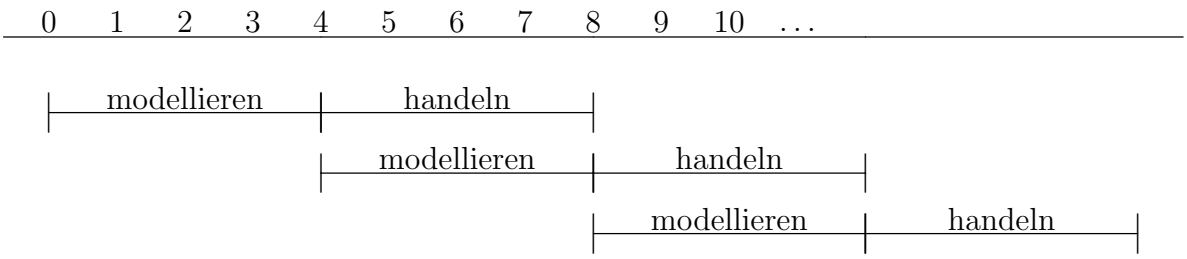
	Engle-Granger		Johansen	
	Preise	Log-Preise	Preise	Log-Preise
1 %	0 %	0 %	0 %	0 %
5 %	0 %	0 %	0 %	0 %
10 %	2.5 %	2.4 %	3.4 %	3.3 %

Zu den Signifikanzniveaus 1 % und 5 % wird die Nullhypothese integrierter - und somit nicht stationärer - Spreads kein einziges mal abgelehnt. Zum 10 % Niveau wird der Spread nur in 2.4 % bis 3.4 % der Fälle als stationär klassifiziert. Das deutet sehr stark darauf

hin, dass die aufgefundenen Kointegrationsbeziehungen hauptsächlich Scheinbeziehungen sind. Sollte ein Aktienpaar tatsächlich kointegriert sein, dann ist diese Beziehung jedoch nicht stabil. Das bedeutet, dass sich die Parameter μ und β zu stark verändern bzw. nach kurzer Zeit bereits keine Kointegrationsbeziehung mehr existiert.

5.2 Parameter für die Handelsstrategie

Die bisherigen Untersuchungen auf stabile Kointegrationsbeziehungen deuten bereits darauf hin, dass die vorliegenden DAX und MDAX Kurse nicht fürs Pairs-Trading basierend auf dem Kointegrations-Ansatz geeignet sind - zumindest nicht in den gegebenen 15-Minuten Intervallen. Dennoch wurden einige Pairs-Trading Strategien implementiert und anhand der historischen Kursdaten getestet. Bei diesen Handelsstrategien handelt es sich im Wesentlichen um die bereits in Abschnitt 4 vorgestellten, wobei die verwendeten Parameter nun erläutert werden:
Wir betrachten Modellierungs- und Handelsphasen von jeweils vier Wochen Länge, welche immer um vier Wochen versetzt werden. Dadurch ergibt sich das folgende Schema:



Die für die Handelsphase notwendigen Parameter werden in der direkt davor liegenden Modellierungsphase geschätzt. Für alle als kointegriert klassifizierte Aktienpaare erhalten wir die Regressionsparameter $\hat{\mu}$ bzw. $\hat{\beta}$ sowie die historische Standardabweichung $\hat{\sigma}$ des Spreads aus der Modellierungsphase. Die verwendeten, Strategie-definierenden Parameter sind nun folgendermaßen gegeben:

- Den Schwellwert th setzen wir gleich der einfachen historischen Standardabweichung $\hat{\sigma}$ des Spreads.
- Die Stop-Loss Schranke $\bar{th}$ setzen wir gleich der dreifachen historischen Standardabweichung $\hat{\sigma}$ des Spreads.
- Die maximale Dauer t_{max} eines Handelszyklus setzen wir gleich 3/4 Wochen.

- Die Hedge-Ratio γ definieren wir - wie in Abschnitt 4.2.2 motiviert - abhängig davon, ob die Preise oder die logarithmierten Preise für die Untersuchung auf Kointegration verwendet werden:

$$\gamma := \begin{cases} \hat{\beta}, & \text{Preise} \\ \hat{\beta}\frac{A_t}{B_t}, & \text{Log-Preise} \end{cases}$$

- Die gehandelten Stückzahlen der Aktien A und B ergeben sich aus der Hedge-Ratio γ[15] und einem gewählten Gesamtumsatz M durch

$$mA_t + m|\gamma|B_t = M$$

- Den Gesamtumsatz M setzen wir gleich 10000 €.

Damit sind für die „normale“ Handelsstrategie alle notwendigen Parameter gegeben: Ein Handelszyklus wird zu einem Zeitpunkt t_0 in der Handelsphase gestartet, falls zu diesem Zeitpunkt keine offenen Positionen existieren (d.h. es kann pro Aktienpaar immer nur ein Handelszyklus gleichzeitig offen sein) und falls der Spread hinreichend weit von seinem Mittelwert entfernt ist:

(1) $A_{t_0} + \hat{\beta}B_{t_0} \geq \hat{\mu} + th$: Leerverkauf von m Stück der Aktie A
Kauf von $m\gamma$ Stück der Aktie B

(2) $A_{t_0} + \hat{\beta}B_{t_0} \leq \hat{\mu} - th$: Kauf von m Stück der Aktie A
Leerverkauf von $m\gamma$ Stück der Aktie B

Ein offener Handelszyklus wird geschlossen, falls eine der folgenden drei Bedingungen erfüllt ist:

(1) Der Spread kehrt zu seinem Mittelwert $\hat{\mu}$ zurück.

(2) Der Spread durchschreitet die Stop-Loss Schranke $\hat{\mu} \pm t\bar{h}$.

(3) Der Handelszyklus ist länger als $t_{max} = 3/4\ Wochen$ offen.

Wir haben bereits gesehen, dass zeitlich konstante Regressions-Parameter μ und β zwar wünschenswert, bei den vorliegenden Kursdaten jedoch nicht realistisch sind. Daher

[15] Die Hedge-Ratio γ ist bei einem negativen Kointegrations-Koeffizienten β ebenfalls negativ. Wird ein Handelszyklus eingegangen, dann sind beide eingegangen Positionen Long- oder Short-Positionen. Daher ist für den Gesamtumsatz der absolute Betrag von γ zu betrachten.

betrachten wir die folgenden drei Varianten, mit möglichen Veränderungen dieser Werte umzugehen:

(U) Die Variante *Update* verläuft analog zum oben vorgestellten „klassischen“ Pairs-Trading. Nach jedem beendeten Handelszyklus werden jedoch die Parameter μ und β sowie die sich daraus ergebende Standardabweichung des historischen Spreads neu bestimmt. Dafür werden die Kursdaten der unmittelbar vorhergehenden vier Wochen verwendet.

(KF) Die Variante *Kalman-Filter* mittels Maximum-Likelihood Schätzung der Tuning-Parameter, d.h. der Varianzen $(\sigma_\eta^2, \sigma_\nu^2, \sigma_\epsilon^2)$ (vgl. Abschnitt 4.2.4). Die Startwerte für $\hat{\mu}_t$ und $\hat{\beta}_t$ stammen von der Schätzung des Regressionsmodelles aus der Modellierungsphase.

(KFm) Die Variante *Kalman-Filter* mittels *manueller* Adaptierung der Tuning-Parameter. Die durch Maximum-Likelihood Schätzung erhaltenen Varianzen $(\sigma_\eta^2, \sigma_\nu^2, \sigma_\epsilon^2)$ werden mit einem Faktor von 10^{-2} verkleinert, um stabilere (d.h. zeitlich weniger variierende) Schätzer für μ_t und β_t zu erhalten.

Zusätzlich zu der Stop-Loss Schranke $\bar{th}$ sollen diese Strategien dazu beitragen, eventuelle Veränderungen der Kointegrationsbeziehung rechtzeitig zu erkennen und so größere Verluste zu vermeiden.

5.2.1 Transaktionskosten

Üblicherweise werden Pairs-Trading Handelsstrategien von Computern automatisiert durchgeführt. Die Transaktionskosten für das Handeln der DAX und MDAX Aktien übernehmen wir deshalb von XETRA®, wobei es sich bei XETRA®um ein elektronisches Handelssystem für an der Frankfurter Börse notierte Wertpapiere[16] handelt. Diese Transaktionskosten sind abhängig vom gehandelten Volumen und vom gewählten Gebührenmodell[17] und bestehen aus bei jedem Handel zu entrichtenden Fixkosten und sogenannten wertbasierten Entgelten. Für die Untersuchungen der Handelsstrategien wurden folgende Werte verwendet:

[16] vgl. http://www.boerse-frankfurt.de/boersenlexikon/Xetra (Stand 01. September 2016)

[17] Es existieren drei Gebühren- oder auch Entgeltmodelle, *High Volume*, *Medium Volume* und *Low Volume*. Diese unterscheiden sich in den monatliche erforderlichen Mindesttransaktionsentgelten und in den Transaktionskosten.
vgl. http://www.xetra.com/xetra-de/handel/handelsinformationen/entgelte-und-gebuehren (Stand 01. September 2016)

(F) Fixkosten: 60 Cent pro Transaktion bis zu einem Betrag von €12500

(V) Variable Kosten: 0.48 Basispunkte (das sind (0.48/100) %) ab einem Betrag von €12500

Hinzukommt das sogenannte *Clearing*, die *„Verrechnung von Forderungen und Verbindlichkeiten aus Wertpapiergeschäften.“* [18] Verwendet wurden hierfür die folgenden Werte:

(F) Fixkosten: 6 Cent pro Transaktion

(V) Variable Kosten: 0.08 Basispunkte (das sind (0.08/100) %), maximal jedoch €4

Dadurch ergeben sich bei einem gehandelten Gesamtumsatz M von €10000 durchschnittliche Transaktionskosten von ca. €2.80 je Handelszyklus.

5.3 Resultate

Die Resultate für die soeben gesprochenen Handelsstrategien unter Berücksichtigung der Transaktionskosten sind in Tabelle 1 zusammengefasst. Die erste Spalte gibt an, welcher Test auf Kointegration verwendet und ob die Preise oder die logarithmierten Preise betrachtet wurden. Weiters wird angegeben, wie viele Paare insgesamt auf allen 48 Zeitfenstern durch diesen Kointegrationstest als kointegriert klassifiziert wurden. Für diese Analyse werden alle als kointegriert klassifizierten Paare gehandelt, d.h. es wird keine weitere Selektion der Aktienpaare durchgeführt, wie dies in Abschnitt 4.1.1 besprochenen wird. Beim Engle-Granger Test ist es außerdem möglich, dass sowohl das Aktienpaar (A, B) - d.h. bei den Aktienkursen A_t handelt es sich um die abhängige Variable in der Regression - als auch das Aktienpaar (B, A) positiv auf Kointegration getestet werden. Auch in diesem Fall werden beide Aktienpaare in der Handelsphase betrachtet. Tatsächlich würde ein Investor sich in diesem Fall für eines der beiden Aktienpaare entscheiden, beispielsweise anhand der in Abschnitt 4.1.1 besprochenen Auswahlkriterien.
Die zweite Spalte gibt an, welche der vier besprochenen Handelsstrategien betrachtet wird: „klassisch“ , „Update“ , „KF“ bzw. „KFm“ .
Anschließend folgen einige deskriptive Statistiken der Gewinne bzw. Verluste dieser Handelsstrategien (Minimum und Maximum, Median und Mittelwert sowie erstes und drittes Quartil).
Nicht jedes sich im Portfolio befindliche Aktienpaar wird tatsächlich gehandelt, beispielsweise falls der Spread in der Handelsphase nie die Schwellwerte $\mu \pm th$ durchschreitet.

[18] vgl. http://www.boerse-frankfurt.de/boersenlexikon/Clearing (Stand 01. September 2016)

Die letzten beiden Spalten geben die Anzahl der tatsächlich gehandelten Paare - das sind jene Paare, bei denen mindestens ein Handelszyklus stattgefunden hat - und die Anzahl der insgesamt durchgeführten Handelszyklen an. Bedingt durch eine hohe Anzahl an als kointegriert klassifizierter Paare, welche jedoch nicht gehandelt werden, ist der Median der Gewinne bzw. Verluste bei fast allen Handelsstrategien gleich Null. Der Mittelwert liegt gemessen am Interquartilabstand sehr nahe bei Null. Werden alle als kointegriert klassifizierten Aktienpaare zum Handeln verwendet, dann erzielt keine der getesteten Handelsstrategien einen signifikant positiven Gewinne. Eine weitere Selektion der Aktienpaare wie in Abschnitt 4.1.1 besprochen erscheint demnach sinnvoll. Die Resultate für weiter selektierte Portfolios werden jedoch nur unwesentlich besser, wie nun für den Fall der „klassischen“ Handelsstrategie des Engle-Granger Tests der Aktienpreise (mit vier Wochen Modellierungs- und 4 Wochen Handelsphase) exemplarisch besprochen wird:
Die erste Grafik zeigt die empirische Verteilung der Anzahl der Handelszyklen.

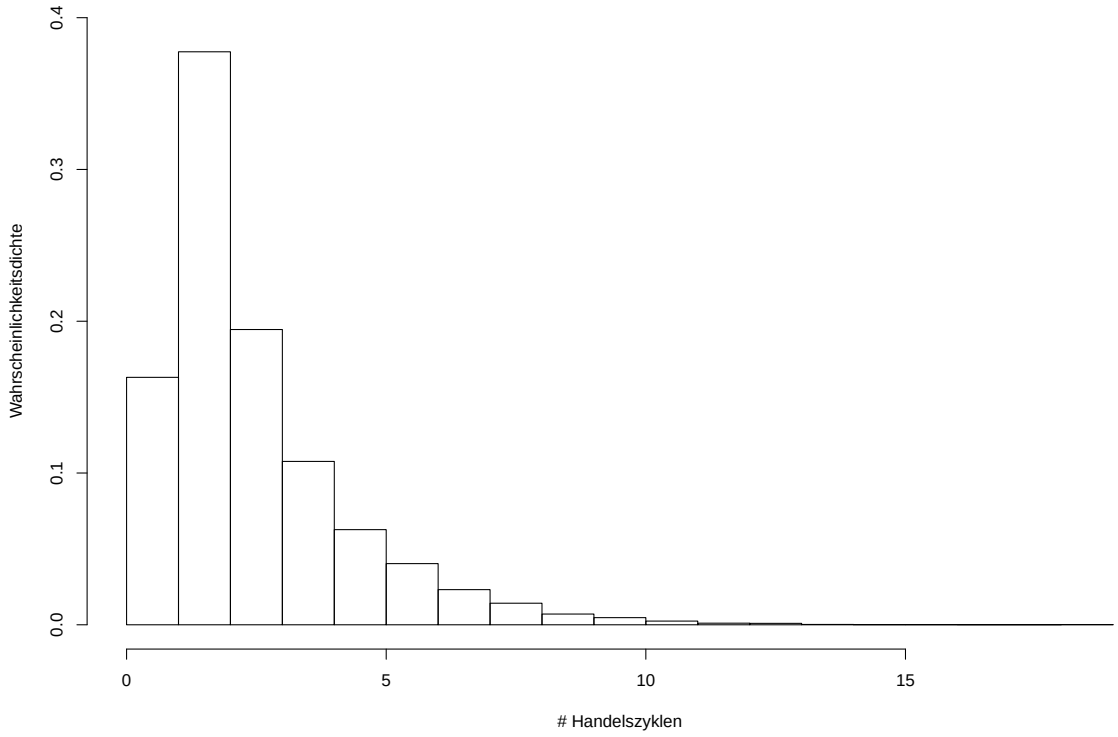


Abbildung 8: Empirische Verteilung der Anzahl der Handelszyklen.

Man erkennt, dass mehr als die Hälfte aller Paare maximal einen Handelszyklus aufweisen. Aktienpaare mit genau Null Handelszyklen machen klarerweise weder Gewinn

		Minimum	1. Quartil	Median	Mittelwert	3. Quartil	Maximum	# gehandelt	# Zyklen
EG: Preise	klassisch	-1878.0	-96.7	-2.9	1.4	46.5	1426.0	15140	35042
	Update	-761.7	-95.9	-0.7	2.5	48.4	1369.0	15140	35938
18089 Paare	KF	-1442.0	-116.6	0.0	1.9	72.1	1679.0	15181	70903
	KFm	-1233.0	-121.5	0.0	6.1	83.5	1596.0	15793	57901
EG: Log-Preise	klassisch	-1863.0	-92.3	0.0	0.6	38.6	1392.0	14417	33041
	Update	-769.7	-91.7	0.0	1.3	38.4	1365.0	14417	33941
18213 Paare	KF	-849.4	-115.0	0.0	1.9	71.0	1847.0	15333	71316
	KFm	-1310.0	-122.8	0.0	6.8	84.6	1791.0	15976	58733
JO: Preise	klassisch	-765.1	-108.3	0.0	-1.5	42.3	1459.0	8046	17625
	Update	-765.1	-106.4	0.0	0.0	46.9	1595.0	8046	18398
9847 Paare	KF	-1529.0	-129.8	0.0	-4.3	65.7	1841.0	8239	36787
	KFm	-1923.0	-135.0	-0.5	0.6	77.5	1220.0	8498	30601
JO: Log-Preise	klassisch	-758.8	-103.1	0.0	-1.1	35.0	1563.0	7725	16777
	Update	-758.8	-100.9	0.0	-0.4	37.6	1625.0	7725	17503
9971 Paare	KF	-795.5	-127.4	0.0	-3.0	68.7	1792.0	8367	37164
	KFm	-1949.0	-135.9	-1.6	2.2	81.9	1838.0	8678	31584

Tabelle 1: Zusammenfassung der Handelsstrategien bei 4 Wochen Modellierungsphase und 4 Wochen Handelsphase.

noch Verlust. Jene Aktienpaare mit genau einem Handelszyklus führen jedoch häufig zu sehr hohen Verlusten, da das Ende des Handelszyklus oft durch das Durchschreiten der Stop-Loss Schranke $t\bar{h}$ bedingt wird. Die Statistiken für die Gewinne bzw. Verluste jener Aktienpaare mit genau einem Handelszyklus sind folgendermaßen gegeben:

Minimum	1. Quartil	Median	Mittelwert	3. Quartil	Maximum
-532.60	-159.50	-115.20	-106.40	-76.33	833.70

Die nächsten Abbildungen zeigen die empirische Verteilung der Gewinne bzw. Verluste, oben für alle als kointegriert klassifizierten Paare (d.h. auch jene Paare ohne Handelszyklen) und unten für alle Paare mit mindestens einem Handelszyklus.

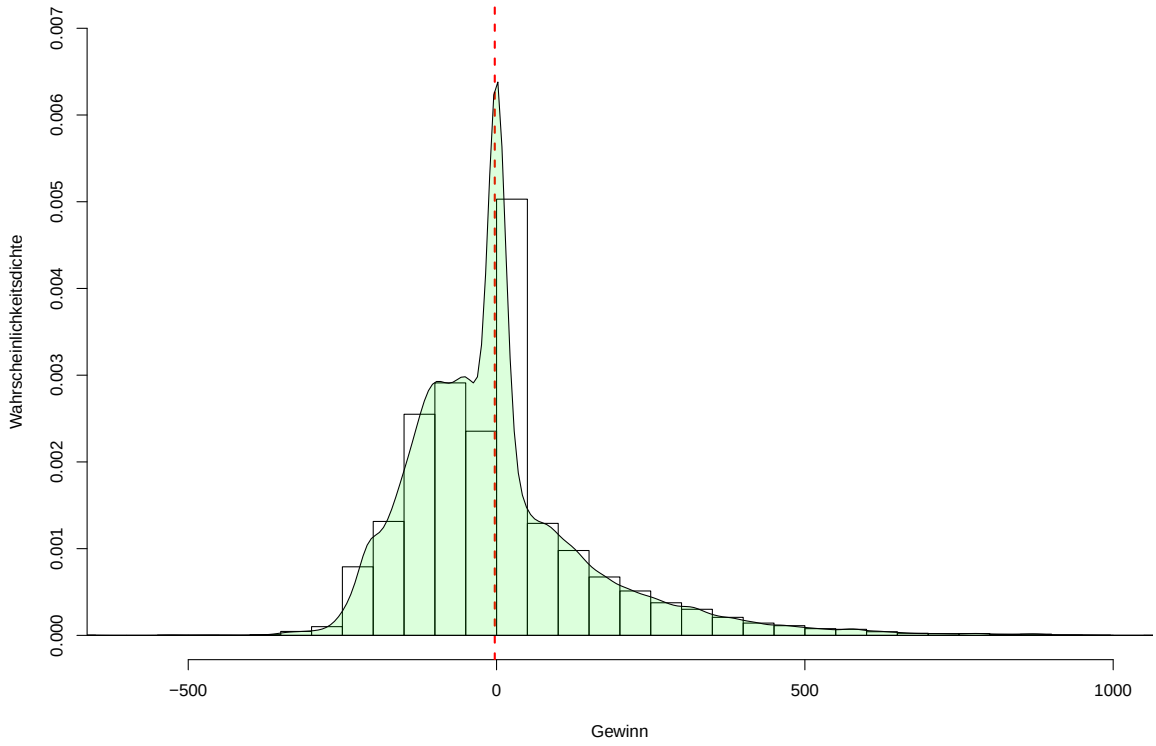


Abbildung 9: Verteilung der Gewinne bzw. Verluste; alle Paare.

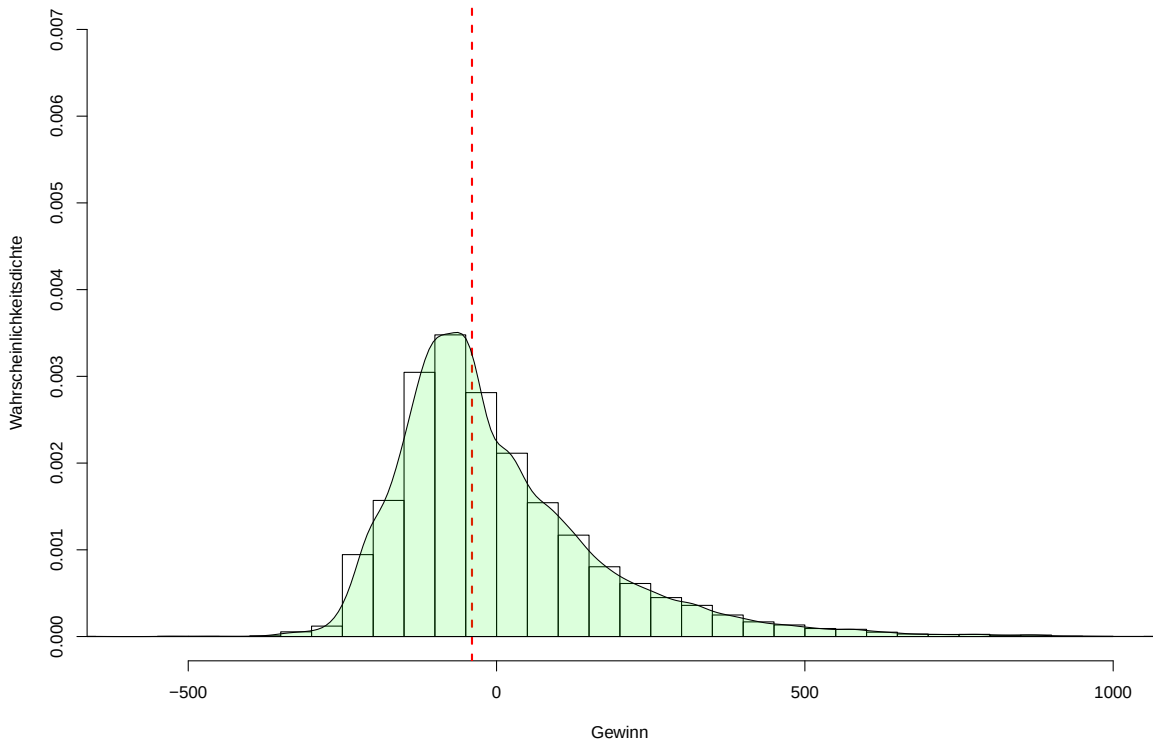


Abbildung 10: Verteilung der Gewinne bzw. Verluste; mindestens ein Handelszyklus.

Die obere Verteilung hat eine Spitze bei Null, da viele Paare nicht gehandelt werden. Die Asymmetrie der beiden Verteilungen folgt aus der Beschränkung der Verluste durch die Stop-Loss Schranken.
Die folgende Grafik zeigt Boxplots der Gewinne bzw. Verluste, wobei die Aktienpaare für einige Auswahlkriterien in Teilmengen entsprechend der Quartile unterteilt wurden. Auch hier lässt sich keine signifikante Verbesserung der Resultate erkennen.

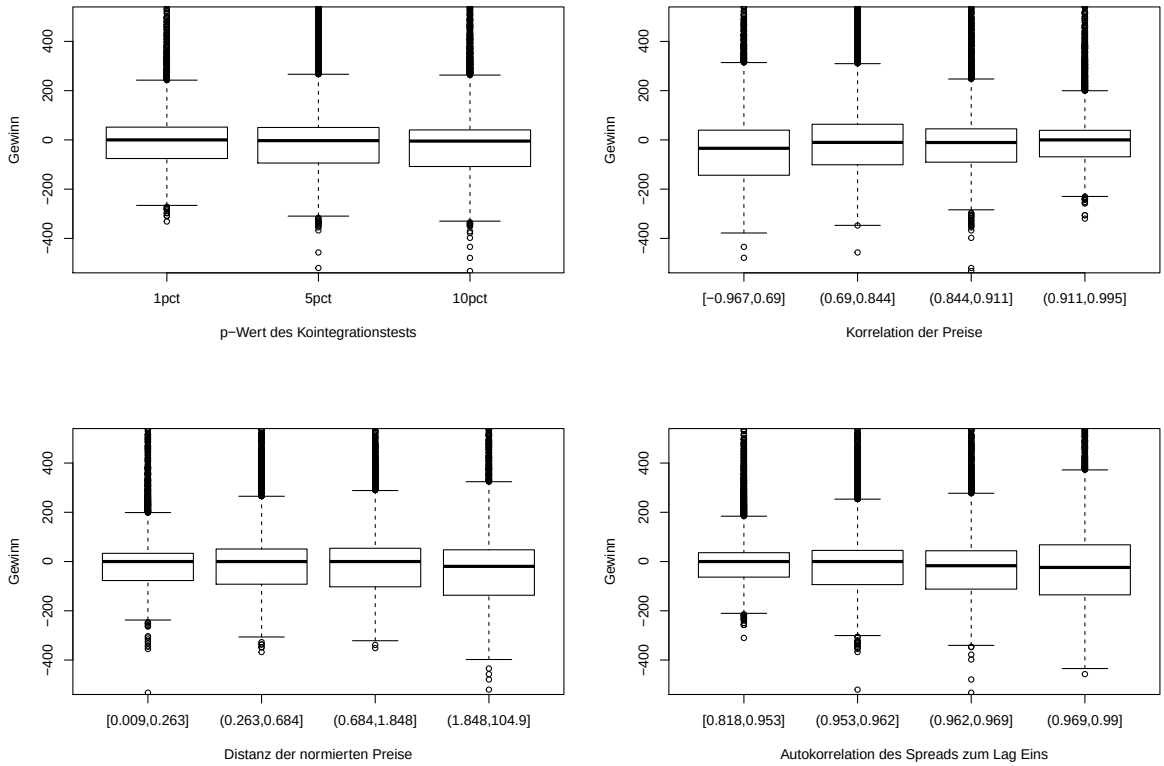


Abbildung 11: Boxplots der Gewinne bzw. Verluste.

Ebenso führt es zu keinen wesentlich besseren Gewinnen, falls die nach dem Engle-Granger Verfahren als kointegriert klassifizierten Aktienpaare ebenfalls mittels Johansen-Test als kointegriert klassifiziert werden. Für die folgenden Boxplots wurden die (mittels des Engle-Granger Verfahrens klassifizierten) Aktienpaare in Teilmengen unterteilt, wobei die Kategorien dem Signifikanz-Niveau des Johansen-Tests entsprechen („NA“ bedeutet, dass diese Aktienpaare nicht durch den Johansen-Test als kointegriert klassifiziert werden).

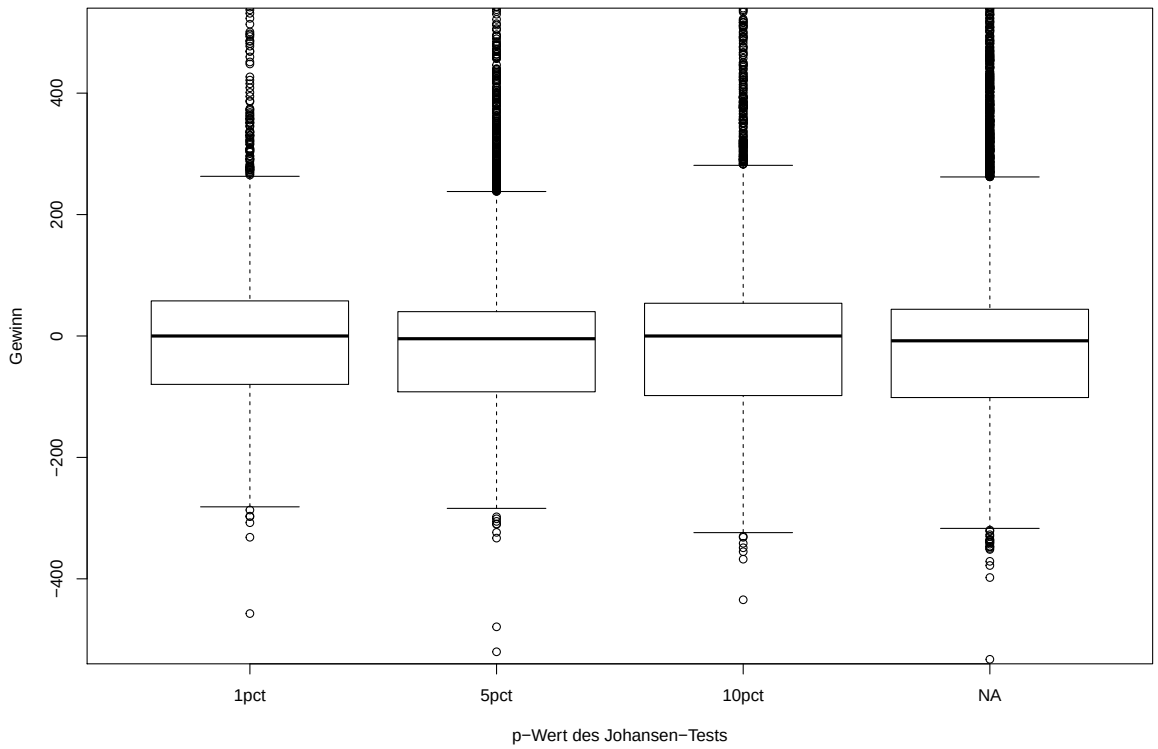


Abbildung 12: Boxplots der Gewinne bzw. Verluste; Kategorien: p-Werte des Johansen-Tests.

Die Ergebnisse der anderen Handelsstrategien sind ähnlich zu den soeben präsentierten. Analoge Resultate erhält man auch, falls unterschiedlich lange Modellierungs- und Handelsphasen verwendet werden. Beispielsweise sind die Resultate beim Engle-Granger Verfahren für die Preisdaten bei einer doppelt so langen Modellierungsphase von acht Wochen und gleichlanger Handelsphase von vier Wochen folgendermaßen gegeben:

	Minimum	1. Quartil	Median	Mittelwert	3. Quartil	Maximum
klassisch	-856.400	-137.500	0.000	3.648	67.270	1376.000
Update	-856.400	-137.200	0.000	4.616	62.580	1278.000
KF	-1934.000	-116.900	0.000	5.595	76.910	1094.000
KFm	-722.200	-119.400	0.000	3.644	70.860	1325.000

6 Fazit

Ziele dieser Diplomarbeit waren eine Überprüfung von insgesamt 61 DAX und MDAX Kursen auf Kointegration und eine Implementation *möglichst* profitabler Handelsstrategien basierend auf Kointegrationsmodellen.
Bereits die Resultate zur Fragestellung, ob überhaupt kointegrierte Aktienpaare gefunden werden können, deuten darauf hin, dass die DAX und MDAX Kurse keine *stabilen* Kointegrationsbeziehungen aufweisen. Werden alle möglichen Aktienpaare auf dem gesamten Zeitraum von Jänner 2012 bis Oktober 2015 auf Kointegration untersucht, wobei der Zeitraum hierfür in 48 direkt aneinander anschließende vierwöchige Zeitfenster unterteilt wurde, dann wird die Nullhypothese H_0: „das Aktienpaar ist nicht kointegriert“ ungefähr zum Prozentsatz des Signifikanzniveaus des verwendeten Kointegrationstests verworfen. Daher liegt die Vermutung nahe, dass es sich hauptsächlich um sogenannte Scheinbeziehungen handelt. Diese Vermutung wird durch weitere Tests auf Stabilität der Kointegrationsbeziehungen weiter bestätigt.
Dennoch wurden vier unterschiedliche Handelsalgorithmen in R implementiert: eine „klassisches“ Strategie bei der die Kointegrations-Parameter konstant bleiben sowie drei „adaptive“ Strategien, bei denen die Parameter sich mit der Zeit verändern dürfen. Alle vier Strategien wurden sowohl mittels des Engle-Granger Kointegrationstests als auch mittels des Johansen-Tests durchgeführt, wobei der Median der Gewinne bzw. Verluste fast immer Null ist. Anpassungen bzgl. der Länge der Modellierungs- bzw. Handelsphase sowie weitere Auswahlkriterien, welche Aktienpaare überhaupt gehandelt werden *dürfen*, haben zu keiner deutlichen Verbesserung der Resultate geführt.
Insgesamt führt dies zum Ergebnis dieser Machbarkeitsstudie, dass die DAX und MDAX Kurse in 15-Minuten Intervallen nicht zum Pairs-Trading basierend auf Kointegrationsbeziehungen geeignet sind. Interessant wäre hier eventuell eine Untersuchung im hochfrequenten Bereich, beispielsweise von 1-Minuten Intervallen oder sogar der sogenannten *Tickdaten*[19]. Falls Kointegrationsbeziehungen der DAX und MDAX Kurse existieren, dann sind diese nur sehr kurzlebiger Natur und müssen schnell genug ausgenutzt werden.

[19] In den Tickdaten werden alle Preisbewegungen des zugehörigen Aktientitels angezeigt. vgl. https://www.wienerborse.at/wissen/begriffsdefinitionen-und-formeln/tickdaten/ (Stand 01. September 2016)

Literaturverzeichnis

[1] Andrade, C. Sandro, Vadim di Pietro, and Mark S. Seasholes, "Understanding the Profitability of Pairs Trading". Working paper, UC Berkeley Haas School, 2005.

[2] Avellaneda Marco and Jeong-Hyun Lee, "Statistical Arbitrage in the U.S. Equities Market", Quantitative Finance Vol. 10 (2010).

[3] Bowen, David, Mark C. Hutchinson, and Niall O'Sullivan, "High-Frequency Equity Pairs Trading: Transaction Costs, Speed of Execution, and Patterns in Returns". The Journal of Trading Vol. 5, no. 3 (2010): pp. 31-38.

[4] Caldeira, João F. and Guilherme V. Moura, "Selection of a Portfolio of Pairs Based on Cointegration: The Brazilian Case", 2012.

[5] Campbell John Y. and Pierre Perron, "Pitfalls and Opportunities: What Macroeconomists Should Know About Unit Roots", NBER Macroeconomics Annual Vol. 6 (1991), Blanchard and Fischer.

[6] Chiu, Mei Choi, and Hoi Ying Wong, "Dynamic Cointegrated Pairs Trading: Time-Consistent Mean-Variance Strategies", Working paper, The Chinese University of Hong Kong, 2012.

[7] Dickey, D. A. and W. A. Fuller, "Likelihood ratio statistics for autoregressive time series with a unit root", Econometrica Vol. 49 (1981): pp. 1057-1072.

[8] Do, Biinh and Robert Faff, and Kais Hamza, "A New Approach to Modeling and Estimation for Pairs Trading", Working paper, Monash University, 2006.

[9] Do, Binh, and Robert Faff. "Does Simple Pairs Trading Still Work?" Financial Analyst Journal Vol. 66, no. 4 (2010): pp. 83-95.

[10] Do, Binh, and Robert Faff. "Are Pairs Trading Profits Robust to Trading Costs?" Journal of Financial Research Vol. 35, no. 2 (2012): pp. 261-287.

[11] Elliot, J. Robert, John van der Hoek, and William P. Malcolm. "Pairs Trading". Quantitative Finance Vol. 5, no. 3 (2005): pp. 271-276.

[12] Engle, Robert F., and C. W. J. Granger. "Co-Integration and Error Correction: Representation, Estimation, and Testing". Econometrica Vol. 55, no. 2 (1987): pp. 251-276.

[13] Föllmer, Hans and Alexander Schied, "Stochastic Finance: An Introduction in Discrete Time", 2nd Edition, De Gruyter Studies in Mathematics 27 (2008), ISBN: 978-3-11-021207-5.

[14] Franke J., W. K. Härdle and C. M. Hafner. "Statistics of Financial Markets: An Introduction". Springer (2008), ISBN: 978-3-540-76272-0.

[15] Gatev, Evan, William M. Goetzmann, and K. Geert Rouwenhorst. "Pairs Trading: performance of a Relative Value Arbitrage Rule". The Review of Financial Studies Vol. 19, no. 3 (2006): pp. 797-827.

[16] Granger, C. W. J. "Some Properties of Time Series Data and Their Use In Econometric Model Specification". Journal of Econometrics Vol. 16, no. 1 (1981): pp. 121-130.

[17] Granger C. W. J. and P. Newbold. "Spurious regression in econometrics". Journal of Econometrics Vol. 2 no. 2 (1974): pp. 111-120.

[18] Hogan, S., R. Jarrow, M. Teo, and M. Warachka. "Testing Market Efficiancy Using Statistical Arbitrage With Applications to Momentum and Value Strategies". Journal of Financial Economics Vol. 73, no. 3 (2004): pp. 525-565.

[19] Johansen, S. "Statistical analysis of cointegrating vectors". Journal of Economic Dynamics and Control Vol. 12 (1988): pp. 231-254.

[20] Johansen, S. "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models". Econometrica, Vol. 59, No. 6 (1991): pp. 1551-1580.

[21] Johansen, S., "Likelihood based Inference in Cointegrated Vector Autoregressive Models". Oxford University Press, Oxford, 1996.

[22] Johansen, S., "Modelling of cointegration in the vector autoregressive model". Economic Modelling Vol. 17 (2000): pp. 359-373.

[23] Johansen, S. and Katarina Juselius. "Maximum Likelihood Estimation and Inference on Cointegration ? with Applications to the Demand for Money", Oxford Bulletin of Economics and Statistics Vol. 52, no. 2 (1990): pp. 169?210.

[24] Kishore, Vayu, "Optimizing Pairs Trading of US Equities in a High Frequency Setting", Wharton Research Scholars Journal, Paper 92, 2012.

[25] Lin, Yan-Xia, Michael McCrae, and Chandra Gulati. “Loss Protection in Pairs Trading Through Minimum Profit Bounds: A Cointegration Approach”. Journal of Applied Mathematics and Decision Sciences (2006): pp. 1-14.

[26] MacKinnon, James G. “Critical Values for Cointegration Tests”. University of California San Diego Discussion Paper 90-4 (1991): pp. 267-276.

[27] MacKinnon, James G. “Critical Values for Cointegration Tests”, Working Papers 1227, Queen’s University, Department of Economics, 2010.

[28] Miao, George J.“High Frequency and Dynamic Pairs Trading Based on Statistical Arbitrage Using a Two-Stage Correlation and Cointegration Approach”, International Journal of Economics and Finance Vol. 6, No. 3 (2014).

[29] Mudchanatongsuk, Supakorn, James A. Primbs, and Wilfred Wong. “Optimal Pairs Trading: A Stochastic Control Approach”. American Controll Conference, Wanshington, (2008).

[30] Phillips, P. “Understanding spurious regression in econometrics”. Journal of Econometrics Vol. 33 no. 3 (1986): pp. 311-340.

[31] Phillips, P. and S. Ouliaris. “Asymptotic properties of residual based tests for cointegration”. Econometrica: Journal of the Econometric Society Vol. 58 no. 1 (1990): pp. 165-193.

[32] Rad, Hossein, Rand Kwong Yew Low, and Robert Faff, “The profitability of pairs trading strategies: Distance, cointegration, and copula Methods”, Working paper, University of Queensland, 2015.

[33] Triantafyllopoulos K. and G. Montana, “Dynamic modeling of mean-reverting spreads for statistical arbitrage”, Computational Management Science Vol. 8 (2009): pp. 23-49.

[34] Vidyamurthy, Ganapathy. “Pairs Trading: Quantitative Methods and Analysis”. New Jersey: John Wiley & Sons, 2007.